**Question 1: Bayesian Networks, Netica (12 + 1 × 4 + 2 × 2 + 3 + 1 × 10 + 7 = 40 marks)**

Expand the Bayes Net you developed in the BN tutorial (available on moodle under the name SmokeAlarm.dne) to include three more events:

**Smoke** (you can see smoke in your apartment),

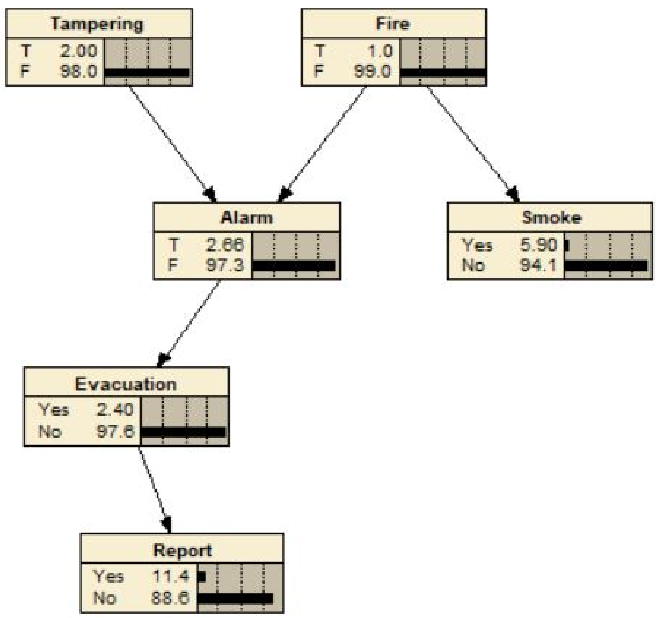
**Evacuation** (your apartment building is evacuated), and

**Report** (the local newspaper writes a report about the evacuation of your apartment).

The probability of smoke when there is fire is 0.95, the probability of smoke when there is no fire is 0.05. When your apartment building has a fire alarm, there is a 0.9 probability that there will be an evacuation, but there is never an evacuation when there is no fire alarm. If there is an evacuation, there is a 0.7 probability that the newspaper will write a report on it, and if there is no evacuation there is a 0.9 probability that the newspaper won't report it.

(a) **Add the necessary nodes and edges to your BN, and input the corresponding conditional probability tables. Justify your expanded network and CPTs. A BN without justification will receive no marks.**

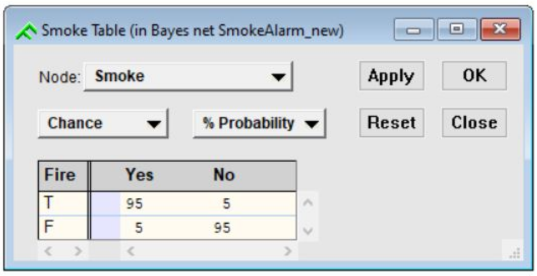
**Answer:**
**Expanded Bayesian Network given as follows:**



**Justification of the Expanded Network:**

**From the given conditions and probabilities, we can justify the following:**

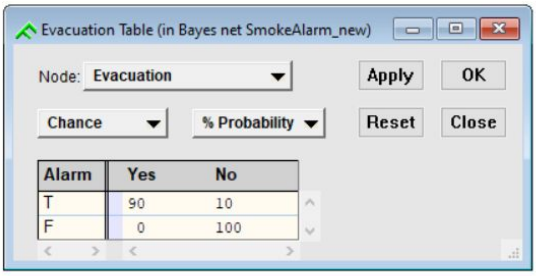
**1) Expanded Smoke Conditional Probability Table :**




| Fire | Yes | No |
|---|---|---|
| T | 95 | 5 |
| F | 5 | 95 |

- When there is **Fire,**
  - Probability of Smoke is **0.95 i.e 95%**
  - Probability of No Smoke is **0.05 i.e 5%**
- When there is **No Fire,**
  - Probability of Smoke is **0.05 i.e 5%**
  - Probability of No Smoke is **0.95 i.e 95%**

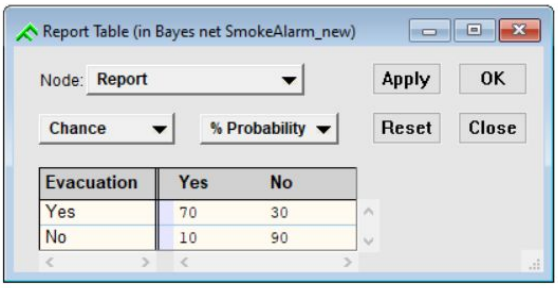
**2) Expanded Evacuation Conditional Probability Table**




| Alarm | Yes | No |
|---|---|---|
| T | 90 | 10 |
| F | 0 | 100 |

- When there is an **Alarm,**
  - Probability of Evacuation is **0.9 i.e 90%**
  - Probability of No Evacuation is **0.1 i.e 10%**
- When there is **No Alarm,**
  - Probability of Evacuation is **0 i.e 0%**
  - Probability of No Evacuation is **1 i.e 100%**

**3) Expanded Report Conditional Probability Table**




| Evacuation | Yes | No |
|---|---|---|
| Yes | 70 | 30 |
| No | 10 | 90 |

- When there is an **Evacuation,**
  - Probability of it being reported is **0.7 i.e 70%**
  - Probability of it not being reported is is **0.3 i.e 30%**

- When there is **No Evacuation,**
    - Probability of it being reported is **0.1 i.e 10%**
    - Probability of it not being reported is is **0.3 i.e 30%**

**(b) Use Netica on the expanded BN to answer the following questions:**

**i. What is the marginal probability that your smoke detector has been tampered with?**

**Answer:**

Marginal Probability that smoke detector has been tampered with -

Looking at the Tampering Event, we get the following -

=> P(Tampering = Yes)

=> 2%

=> **0.02**

**ii. What is the marginal probability that there will be a news report tomorrow?**

**Answer:**

Marginal Probability that there will be a news report tomorrow -

Looking at the Report Event, we get the following -

=> P(Report = Yes)

=> 11.4%

=> **0.114**

**iii. Let's assume that you have observed that there is smoke in your apartment. What is the posterior probability that there will be a news report tomorrow?**

**Answer:**

Here**,**

Smoke = Yes

Report = Yes

Posterior Probability can be given as:

=> P(Report = Yes | Smoke = Yes)

=> **19.3 %**

=> **0.193**

**iv. Let's assume that you have observed that there was no fire, and that there was a news report about your apartment. What is the posterior probability that your smoke detector has been tampered with?**

**Answer:**

**Given**:

Fire = No

Report = Yes

**To Find:**

Tampering = Yes

**Solution:**

Posterior Probability can be given as:

=> P(Tampering = Yes | Report = True, Fire = No)

=> **10.2%**
=> **0.102**

### v. Let's assume that you have observed that there is no smoke in your apartment. What is the posterior probability that your smoke detector has been tampered with? What conditional independence property could help you here?

### Answer:

**Given**:
Smoke = No
**To Find:**
Tampering = Yes
**Solution:**
**Posterior Probability that smoke detector has been tampered with :**
=> P(Tampering = Yes | Smoke = No)
=> **2%**
=> **0.02**
**Conditional independence property could help :**

- Tampering with the Alarm and Generation of Smoke are Independent events.
- Hence, the Tampering Probability will not be affected when we modify the Smoke Values.

### vi. Let's assume that you have observed that there has been a news report about your apartment, and there is no smoke in your apartment. What is the posterior probability that your smoke detector has been tampered with? Given that the news report was observed, why does observing the absence of smoke affect your belief of whether or not your smoke alarm was tampered with?

### Answer:

**Given**:
Report = Yes
Smoke = No
**To Find:**
Tampering = Yes
**Solution:**
**Posterior Probability that smoke detector has been tampered with :**
=> P(Tampering = Yes | Smoke = No, Report = Yes)
=> **10.2%**
=> **0.102**
**Belief on Tampering on the Smoke Alarm in the absence of Smoke:**

- We know that **Smoke** is **DEPENDENT** on **Fire**
- Report is affected by Fire and Tampering via Alarm and Evacuation.
- If there was a Report,
    - And we observe Smoke
    - We can be sure that there was no Tampering.
    - Because :
        - There was Smoke
        - Which led to Alarm ( **Not Tampered )**
        - Which led to Evacuation

        - Which finally led to Report
    - However, here though there was a Report
        - There was no Smoke
        - There was an Alarm ( **As Alarm was Tampered )**
        - There was an Evacuation
        - And, Finally there was a Report
- Clearly, the Alarm was Tampered.

**vii. Let's assume that you have observed that there was no fire, that there was a news report about your apartment, and that there is smoke in your apartment. What is the posterior probability that your smoke detector has been tampered with? How does observing whether or not there is smoke affect your belief of whether or not your smoke detector has been tampered with? Why?**
**Answer:**
**Given**:
Fire = No
Report = Yes
Smoke = Yes
**To Find:**
Tampering = Yes
**Solution:**
**Posterior Probability that smoke detector has been tampered with :**
=> P(Tampering = Yes | Smoke = Yes, Report = Yes, Fire = No)
=> **10.2%**
=> **0.102**
**Belief on Tampering on the Smoke Alarm in the absence of Smoke:**
- Clearly, Smoke is DEPENDENT on Fire.
- And Alarm is DEPENDENT on Fire.
- Alarm is NOT DEPENDENT on Smoke directly.
- Here, it is given that there was No Fire,
    - Hence, Smoke does not have any effect on the Alarm.

**(c) Hypothesize the (conditional) independence properties** of the statements below. Use Netica to check them, and state whether they are true or false. Briefly explain your answers. Answers without explanations will receive no marks.

Note:The graph structure informs us about dependencies between variables, but there may be additional dependences based on the values of the conditional probability tables.

**ANSWER:**
**Three factors which affect Conditional Independence (CI) :**
- **Causal Chain** : Evidence indicates - CI
- **Common Cause** : Evidence indicates - CI
- **Common Effect** : No Evidence indicates - CI

- **Tampering ⊥⊥ Evacuation**
    - **Conditional Independence = False**
    - **Explanation :**
        - Causal Chain :
            - Tampering and Evaluation have a common path with Alarm
            - This makes them a Causal Chain
            - However, there is **NO EVIDENCE** for Alarm
            - Making the path is active
            - Hence, Tampering and Evacuation are **NOT INDEPENDENT**

- **Tampering ⊥⊥ Evacuation | Alarm**
    - **Conditional Independence = True**
    - **Explanation :**
        - Causal Chain :
            - Tampering and Evaluation have a common path with Alarm
            - This makes them a Causal Chain
            - However, there is **EVIDENCE** for Alarm
            - Making the path is Inactive
            - Hence, Tampering and Evacuation are **INDEPENDENT**

- **Tampering ⊥⊥ Evacuation | Smoke**
    - **Conditional Independence = False**
    - **Explanation :**
        - Causal Chain :
            - Tampering and Evacuation have a common path with Alarm
            - However, they do not have Smoke in their path
            - Now, there is **NO EVIDENCE** for Alarm
            - Making the path is active
            - Hence, Tampering and Evacuation are **NOT INDEPENDENT**

- **Tampering ⊥⊥ Fire**
    - **Conditional Independence = True**
    - **Explanation :**
        - Common Effect :
            - Tampering and Fire are responsible for Alarm
            - Alarm is their Common Effect
            - Now, there is **NO EVIDENCE** for Alarm
            - Hence, Tampering and Evacuation are **INDEPENDENT**

- **Tampering ⊥⊥ Fire | Alarm**
    - **Conditional Independence = False**
    - **Explanation :**
        - Common Effect :
            - Tampering and Fire are responsible for Alarm
            - Alarm is their Common Effect
            - Now, there is **EVIDENCE** for Alarm
            - Hence, Tampering and Evacuation are **NOT INDEPENDENT**
- **Alarm ⊥⊥ Smoke**

    - **Conditional Independence = False**
    - **Explanation :**
        - Common Cause :
            - Alarm and Smoke have Fire as their Common Cause
            - And, There is **NO EVIDENCE** for Fire
            - Hence, Alarm and Smoke are **NOT INDEPENDENT**

- **Smoke ⊥⊥ Report**
    - **Conditional Independence = False**
    - **Explanation :**
        - Common Cause and Causal Chain
            - Report is connected to Smoke via Evacuation, Alarm and Fire.
            - Fire is the Common Cause between Smoke and Report
                - It has **NO EVIDENCE**
            - Alarm and Evacuation form a Causal Chain between Smoke and Report
                - They have **NO EVIDENCE**
            - Hence, Report and Smoke are **NOT DEPENDENT**

- **Smoke ⊥⊥ Tampering**
    - **Conditional Independence = True**
    - **Explanation :**
        - Common Effect :
            - Smoke and Tampering are connected via Fire and Alarm
            - Fire is the Common Cause
                - It has **NO EVIDENCE**
            - However, Alarm and it's descendents are the Common Effect
                - They have **NO EVIDENCE**
            - Hence, Smoke and Tampering are **INDEPENDENT**

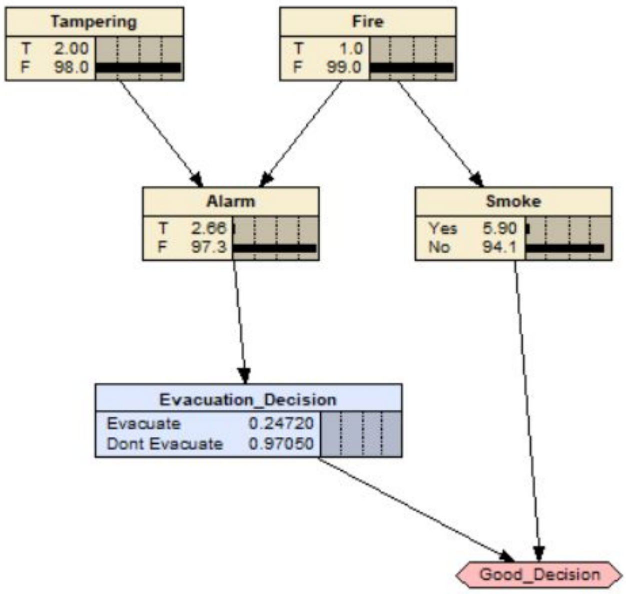
- **Smoke ⊥⊥ Tampering | Alarm**
    - **Conditional Independence = False**
    - **Explanation :**
        - Common Effect :
            - Smoke and Tampering are connected via Fire and Alarm
            - Alarm and it's descendents are the Common Effect
                - Here, Alarm has **EVIDENCE**
            - Hence, Smoke and Tampering are **NOT INDEPENDENT**

- **Smoke ⊥⊥ Tampering | Report**
    - **Conditional Independence = False**
    - **Explanation :**
        - Common Effect :
            - Smoke and Tampering are connected via Fire and Alarm and Alarm's Descendants i.e Evacuation and Report
            - Alarm and it's descendents are the Common Effect
                - Here, Report has **EVIDENCE**
            - Hence, Smoke and Tampering are **NOT INDEPENDENT**

(d) Based on your BN, construct a Bayesian Decision Network (BDN) that decides whether the building should be evacuated. That is, instead of having an Evacuation chance node, you should have a **decision node** that determines whether you should evacuate the building. Specify and justify the information links and the **values in the utility node**. BDNs without justifications will receive no marks.

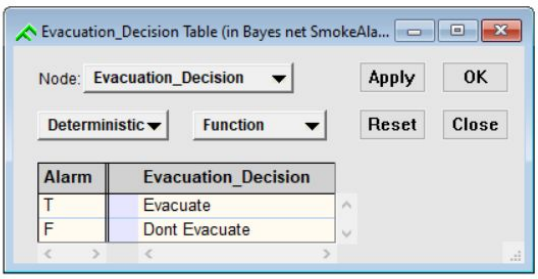
**ANSWER:**

**Bayesian Decision Network (BDN) :**




**Justification**

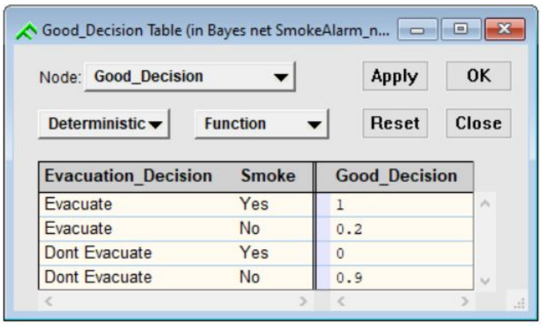
- In my Network :
    - **Evacuation_Decision** : A Decision Node is connected to the Alarm Node and is dependent on the Alarm Node.
    - **Good_Decision** : A Utility Node is connected to the Evacuation_Decision and Smoke Node and is dependent on both these nodes.
- Here, my "**decision node"** is **Evacuation_Decision.**




    - This Decision Node has two values :

- Evacuate
    - When Alarm is True
    - Decision Made is Evacuate
- Don’t Evacuate
    - When Alarm is False
    - Decision Made is Don’t Evacuate

- And my **“Utility node”** is **Good_Decision**




| Evacuation_Decision | Smoke | Good_Decision |
|---|---|---|
| Evacuate | Yes | 1 |
| Evacuate | No | 0.2 |
| Dont Evacuate | Yes | 0 |
| Dont Evacuate | No | 0.9 |

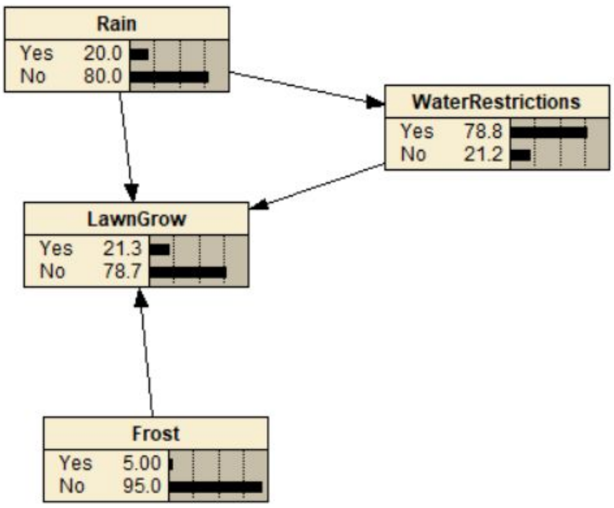
- My Utility Node has four values of Good_Decision
- When we Evacuate and there was smoke,
    - It’s a good decision!
    - Hence, it’s value is 1
- When we Evacuate and there was no smoke,
    - It’s not that good a decision!
    - But, Evacuation was needed as a precautionary measure
    - And, the Evacuation procedure was followed as a protocol keeping in mind the safety of the people
    - Hence, I have assigned a very small value of 0.2 to Good_Decision
- When we Don’t Evacuate and there was smoke,
    - It’s not a good decision at all!
    - When there is Fire, there is 95% chance that there will be smoke
    - Thus, the probability of Fire, given there was smoke is quite high
    - Hence, if there smoke, an Evacuation is a must, keeping in mind the safety and security of people
    - Hence, I have assigned a value of 9 to Good_Decision
- When we Don’t Evacuate and there was no smoke,
    - Now, this is again a good decision!
    - When there was no smoke, there is 95% chance that there was no fire
    - However, we need to keep in mind that sometimes a fire can arise without smoke.
    - Hence, I have assigned a value of 0.9  to Good_Decision

**Question 2: Bayesian Networks, Netica (17 + 4 + 12 + 7 = 40 marks)**

It is coming to the end of winter and Ron is trying to model the factors that affect the state of his lawn. The lawn is currently looking pretty sad, as his children spent all last summer playing backyard cricket, and have worn several bare patches. **However, the area has been in drought for the previous 12 months**. If there is no rain before summer, it will be very hard to get the new lawn to grow, and Ron will waste a lot of time and money. Furthermore, if there is no rain, the authorities could increase the level of water restrictions, meaning that Ron will be unable to water his lawn at all. This would make the chances of his lawn surviving very small indeed. To further complicate the matter, there is a small chance that the area could experience another frost before the weather warms up, which also could damage the new lawn.

**(a) Design a BN using the nodes: Rain, LawnGrow, WaterRestrictions and Frost. Justify your design. A BN without justification will receive no marks.**

**ANSWER :**




**Design Justification:**

- Here, I have assumed 3 contributing factors to the growth of the lawn.
    - Rain
    - Water Restrictions
    - Frost
- Also, Rain will have an impact on the level of Water Restrictions.
    - Hence, there is a link between Rain and Water Restriction.
- In my assumption and understanding :-
    - Rain will have the greatest impact on the growth of lawn
    - Followed by Water Restrictions

    - Finally, Frost will have the least impact on the growth of the lawn

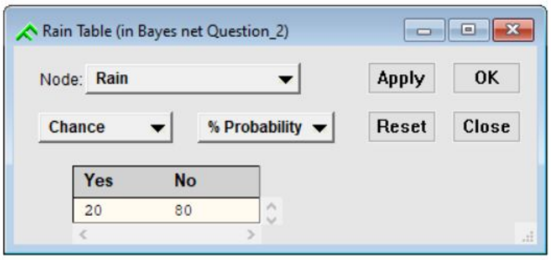
### (b) Inspect your BN and report on any value assignments that will cause d-separation between any sets of nodes. Explain why this is the case. Value assignments without explanations will receive no marks.

- Value Assignments that will cause d-separation:
- When 2 nodes are d- separated by evidence, X and Y are conditionally independent.
- **Frost and WaterRestrictions are d separated by LawnGrow.**
    - LawnGrow is their Common Effect
    - If there was evidence of LawnGrow,
        - Frost and WaterRestrictions will not be INDEPENDENT.
    - If there was NO EVIDENCE of LawnGrow,
        - Frost and WaterRestrictions will be INDEPENDENT.
- **Frost is separated from Rain through LawnGrow.**
    - LawnGrow is their Common Effect
    - If there was evidence of LawnGrow,
        - Frost and WaterRestrictions will not be INDEPENDENT.
    - If there was NO EVIDENCE of LawnGrow,
        - Frost and WaterRestrictions will be INDEPENDENT.

### (c) Quantify the relationships in the network by adding numbers for the CPTs. Justify the numbers in your CPTs. CPTs without justification will receive no marks.

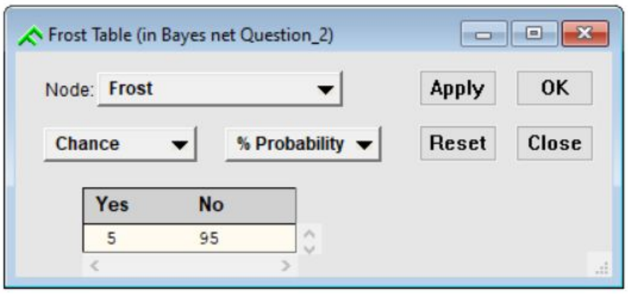
**<u>Justification of the Expanded Network:</u>**
**From the given conditions and probabilities, we can justify the following:**

**Expanded Rain Probability Table :**



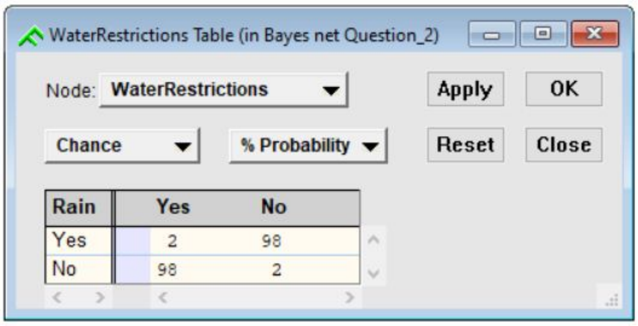

- **Justification:**
    - Since it is given that the area has been in drought for the previous 12 months,
    - I have assumed Probability of Rain would be less than Probability of No Rain
    - Hence, the following:
        - Probability of Rain before Summer is **0.2 i.e 20%**
        - Probability of No Rain before Summer is **0.8 i.e 80%**

**Expanded Frost Probability Table :**




- **Justification:**
    - Since it is given that the area here is a small chance that the area could experience another frost
    - I have assumed Probability of Frost would be less than Probability of No Frost
    - Hence, the following:
        - Probability of Frost is **0.05 i.e 5%**
        - Probability of No Frost is **0.95 i.e 95%**

**Expanded WaterRestrictions Conditional Probability Table :**




- **Justification:**
    - When there is **Rain,**
        - Probability of Increase in Water Restrictions is **0.02 i.e 2%**
        - Probability of No Increase in Water Restrictions is **0.98 i.e 98%**
    - When there is **No Rain,**
        - Probability of Increase in Water Restrictions is **0.98 i.e 98%**
        - Probability of No Increase in Water Restrictions is **0.02 i.e 2%**
- Here, when it rains, my Probability of Increase in Water Restrictions is a very small value.
    - Reason: My assumption is that, if it rains, there will not be any dearth of water and hence there will not be any water restrictions.
    - Similarly, when there is no rain, there will be death of water and there will be enforced water restrictions.

**Expanded LawnGrow Conditional Probability Table :**

LawnGrow Table (in Bayes net Question_2)

| Rain | WaterRestrictions | Frost | Yes | No |
|---|---|---|---|---|
| Yes | Yes | Yes | 70 | 30 |
| Yes | Yes | No | 75 | 25 |
| Yes | No | Yes | 95 | 5 |
| Yes | No | No | 99 | 1 |
| No | Yes | Yes | 1 | 99 |
| No | Yes | No | 2 | 98 |
| No | No | Yes | 5 | 95 |
| No | No | No | 8 | 92 |

**Justification:**

- When it Rains :
    - When there are Water Restrictions :
        - When there is Frost,
            - Probability of Lawn Growth : 0.7 i.e 70%
            - Probability of No Lawn Growth : 0.3 i.e 30%
        - When there is No Frost,
            - Probability of Lawn Growth : 0.75 i.e 75%
            - Probability of No Lawn Growth : 0.25 i.e 25%
    - When there are No Water Restrictions :
        - When there is Frost,
            - Probability of Lawn Growth : 0.95 i.e 95%
            - Probability of No Lawn Growth : 0.05 i.e 5%
        - When there is No Frost,
            - Probability of Lawn Growth : 0.99 i.e 99%
            - Probability of No Lawn Growth : 0.01 i.e 1%
- When it does not Rain :
    - When there are Water Restrictions :
        - When there is Frost,
            - Probability of Lawn Growth : 0.01 i.e 1%
            - Probability of No Lawn Growth : 0.99 i.e %
        - When there is No Frost,
            - Probability of Lawn Growth : 0.02 i.e 2%
            - Probability of No Lawn Growth : 0.98 i.e 98%
    - When there are No Water Restrictions :
        - When there is Frost,
            - Probability of Lawn Growth : 0.05 i.e 5%
            - Probability of No Lawn Growth : 0.95 i.e 95%
        - When there is No Frost,
            - Probability of Lawn Growth : 0.08 i.e 8%
            - Probability of No Lawn Growth : 0.92 i.e 92%

**(d) Using Netica, demonstrate the workings of your BN by determining the probability of the lawn growing in the following cases.**

**1. There is no evidence.**

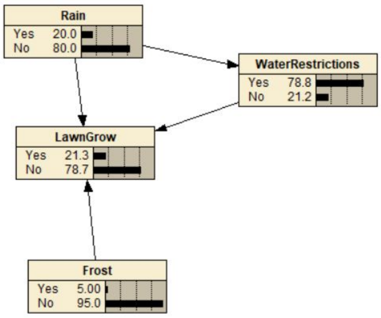
Since, there is no evidence

P(LawnGrow = Yes)

=> Marginal Probability

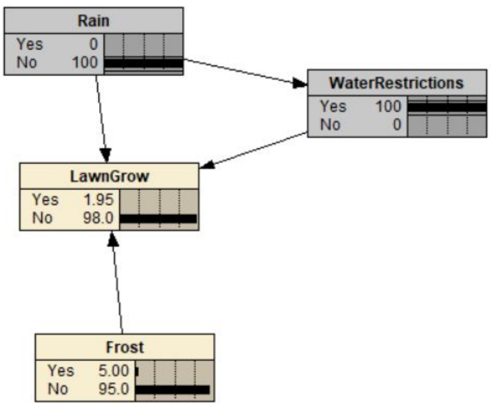
=> 0.213

**Results from Netica:**




**2. There is no rain, and water restrictions have been applied. Explain your results compared to item 1.**

**Results from Netica:**




P(LawnGrow = Yes | Rain = No, WaterRestrictions = Yes)

=> 0.0195

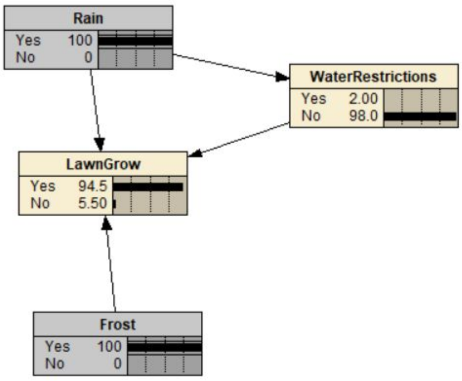
**Explanation:**

- Clearly, we can see the impact of No Rain and Water Restrictions on the growth of the lawn.

- No Rain and Water Restrictions have drastically reduced the Probability of Lawn Growth from 0.213 to 0.0195.
- This is because, in my Model I have assumed Rain to be the most important factor that will affect the growth of lawn.

### 3. There is frost, but it has rained. Explain your results compared to item 1.

**Results from Netica:**




P(LawnGrow  = Yes | Frost = Yes, Rain = Yes )

=> 0.945 %

**Explanation:**

- Clearly, we can see the impact of Rain and No Frost on the growth of the lawn.
- Rain and No Frost have drastically increased the Probability of Lawn Growth from 0.213 to 0.945.
- This is because, in my model, Rain has the most impact on the growth of lawn and Frost does not have that big of an impact on the growth of lawn.

**<u>Question 3: D − separation (9 + 5 + 6 = 20 marks)</u>**

Consider the following Bayesian Network called rental2.dne (available on moodle).

### (a) List the conditions under which you will be able to propagate evidence from Income to Rent charged. That is, which nodes need to be instantiated or uninstantiated so that evidence can be propagated from Income to Rent charged. Explain why this is the case.

### Answer :

- For the path to become active, Nodes need to be **instantiated** are :
  - Property_area
  - Tenant or Happiness
- **Explanation :**
  - When we observe the above network, we can see that there is one path from **Income to Rent.**
  - **Path -** Income .. Property_area .. Housing_prices .. Tenant .. Rent_charged
  - **Path can be EXPANDED as :**
    - **Income -> Property_area <- Housing_prices**
      - If Property_area is observed, path is active
      - If Property_area is not observed, path is inactive
    - **Property_area <- Housing_prices -> Tenant**
      - If Housing_prices is observed, path is active
      - If Housing_prices is not observed, path is in-active
    - **Housing_prices -> Tenant <- Rent_charged**
      - If Tenant is observed, path is active
      - If Tenant is not observed, path is inactive

### (b) Repeat question (a) for propagating evidence from Happiness to Property area (with explanations).

### Answer :

- Here, For the path to become active, No Nodes need to be **instantiated.**
- **Explanation :**
  - Here, we have a Causal Chain :
    - Happiness <- Tenant <- Housing_prices
      - If Tenant is observed, path is Inactive
      - If Tenant is not observed, path is active
  - Here, we have a Common Cause :
    - Tenant <- Housing_prices -> Property_area
      - If Housing_prices is observed, path is Inactive
      - If Housing_prices is not observed, path is active
  - Hence, the evidence from Happiness to Property propagates directly.

**(c) Repeat the above questions under the assumption that there is also an arc from Property area to Tenant (the corresponding BN, rental3.dne, is available on moodle). Compare your results with those obtained above.**

**Answer:**

- **Question a) : Income to Rent**
  - Nodes to be instantiated:
    - Tenant
  - **Explanation:**
    - Path :
      - Income -> Property_area <- Housing_price -> tenant <- Rent_charged
    - Also, we have another Path :
      - Income -> Property_area -> Tenant <- Rent_charged
      - Income -> Property_area -> Tenant  (Causal Chain)
        - Here, when Property_area is not observed, path is active
        - Here, when Property_area is observed, path is active
- **Question b) Happiness to Property Area**
  - No Nodes need to be instantiated
  - **Explanation:**
    - Because of the new link created,
    - There is another **Back Propagation Channel** from **Happiness to Property_Area**

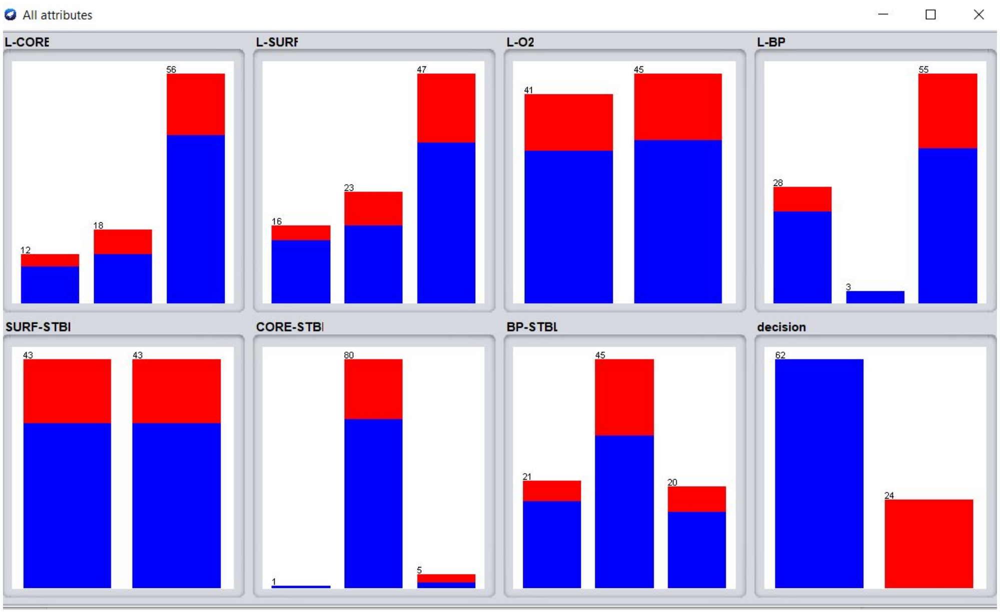
## Question 4: Classification, Decision Trees, Na¨ıve Bayes, k-NN, Weka (40 marks)

Consider the dataset postoperative-patient-data simplified.arff available on moodle. This dataset contains health-status attributes of post-operative patients in a hospital, with the target class being whether the patients should be discharged (S) or remain in the hospital (A). Additional documentation regarding these attributes appears in the arff file.

**1. Before you run the classifiers, use the weka visualization tool to analyze the data, and report briefly on the types of the different variables and on the variables that appear to be important. (4 marks)**

**ANSWER:**

**Weka Visualisation tools to analyze the data:**




**Types of Variables:**

- Here, the Target Variable is **DECISION**
- The **RED** colour - indicates No
- The **BLUE** colour - indicates Yes
- Here, the Target Variable is **DECISION**
- There are 7 variables which determine the Target Variable Decision.

**Important Variables :**

- The variables which help in discriminating the Target Variable the most are important and significant variables.
- Here, **Important Variables are :**
  - CORE-STBL
  - L-BP
  - **Reason :** Here, they clearly **DISCRIMINATES** between the **TWO** values of the Target Variable - **decision.**

- ○ In both their visualisations, we can clearly see the **NO** value of the target variable separated from the **YES** value.

**2.** Run J48 (=C4.5, decision tree), Na¨ıve Bayes and IBk (k-NN) to learn a model that predicts whether a patient should be discharged. Perform 10-fold cross validation, and analyze the results obtained by these algorithms as follows.

Note: Click on the “Choose” bar to select relevant parameters. Explanations of parameters you should try appear below. You should report on performance of at least two variations of the operational parameters, e.g., **minNumObj and unpruned for J48, and KNN and distanceWeighting for k-NN** (the parameters debug and saveInstanceData are not operational).

J48

• binarySplits: whether you use binary splits on nominal attributes when building the trees.

• minNumObj: the minimum number of instances per leaf.

• unpruned: whether pruning is performed (try TRUE and FALSE).

• debug: if set to TRUE, the classifier may output additional information.

• saveInstanceData: whether to save the training data for visualization.

Na¨ıve Bayes (parameter variations are not relevant to this lab)

k-NN (IBk) (under lazy in WEKA)

• KNN: the number of neighbours to use.

• crossValidate: whether leave-one-out X-validation will be used to select the best k value between 1 and the value specified in the KNN parameter.

• distanceWeighting: specifies the distance weighting method used (when k ¿ 1).

• debug: if set to TRUE, the classifier may output additional information.

**(a) J48 (=C4.5) (3 + 2 = 5 marks)**

**i. Examine the decision tree and indicate which are the main variables.**

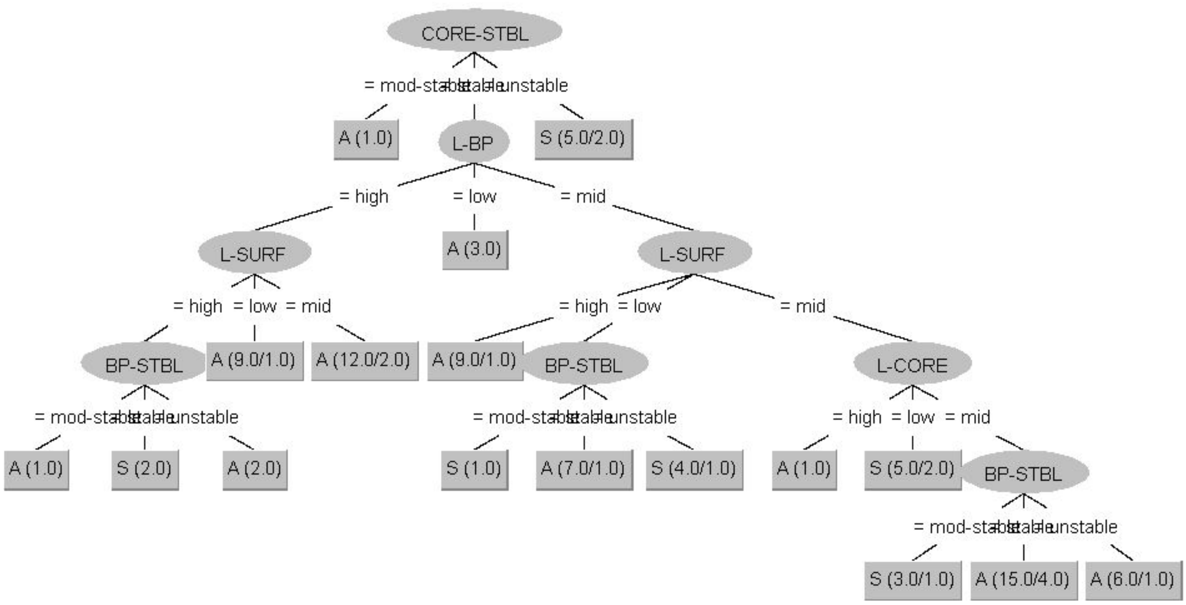
**ANSWER:**

**Decision Tree :**

- In order to get the Decision Tree, we set the following parameters :

**Variations of the operational parameters :**

- binarySplits: False
- **minNumObj:**
    - the minimum number of instances per leaf.
    - Here, it is taken as 2.
    - This is the number of instances at each node which are required for the decision to be evaluated.
- **unpruned:**
    - whether pruning is performed (try TRUE and FALSE).
    - The value was False by default. That time, we could not see the entire tree.
    - When this value was set to True, we could visualise the entire decision tree.
- debug:
    - if set to TRUE, the classifier may output additional information.
- saveInstanceData:
    - whether to save the training data for visualization.
    - Here, we have set it to False.



**Main Variables of the Decision Tree :**

- **CORE - STBL**
- **L- BP**
- **L-SURF**
- Here, the Decision Tree starts from **CORE - STBL.**
- Hence, these are the main variables.

## ii. What is the accuracy of the decision tree? Explain the results in the confusion matrix.

## ANSWER:

**Accuracy :**

- Here, the accuracy of the decision tree is **69.7674%**
- This is taken from “**Correctly Classified Instances”**

```
Number of Leaves  :     17

Size of the tree :      25


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          60               69.7674 %
Incorrectly Classified Instances        26               30.2326 %
```

- **Confusion Matrix :**

```
=== Confusion Matrix ===

  a  b   <-- classified as
 54  8 |  a = A
 18  6 |  b = S
```

  - **Results of the Confusion Matrix :**
    - J48 was able to correctly classify 54 instances of A
    - Also, it incorrectly classified 8 instances of A as B
    - Similarly, it was able to correctly classify 6 instances of B
    - Finally, it incorrectly classified 18 instanced of B as A

**(b) Naive Bayes (2 + 8 + 3 + 2 = 15 marks)**

**i. Explain the meaning of the "probability distributions" in the output, illustrating it with reference to the BP STBL attribute.**

**ANSWER:**

- Probability Distribution is the frequency distribution of the various values of the attribute.
- This classification is in accordance with the algorithm
- Here, in the BP STBL attribute, we have three values:
    - mod-stable
    - stable
    - unstable.

**ii. Calculate (by hand) the probability that a person with the following attribute values would be discharged.**

**L-CORE = mid**

**L-SURF = low**

**L-O2 = good**

**L-BP = high**

**SURF-STBL = stable**

**CORE-STBL = stable**

**BP-STBL = mod-stable**

ii) $P(S|X) = \frac{P(X|S)\,P(S)}{P(X)}$ $\qquad \alpha = \frac{1}{P(X)}$

$P(S|X) = \alpha\, P(X|S)\, P(S)$

Here $X = \begin{cases} \text{L-CORE} = \text{mid} \\ \text{L-SURF} = \text{low} \\ \text{L-O2} = \text{good} \\ \text{L-BP} = \text{high} \\ \text{SURF-STBL} = \text{stable} \\ \text{CORE-STBL} = \text{stable} \\ \text{BP-STBL} = \text{mod-stable} \end{cases}$

A : Remain in the hospital.
S : Discharged.

$\Rightarrow P(S|X) = \alpha\,[P(\text{mid}|S) \cdot P(\text{low}|S) \cdot P(\text{good}|S) \cdot P(\text{high}|S) \cdot P(\text{stable}|S) \cdot P(\text{stable}|S) \cdot P(\text{mod-stable}|S)]\, P(S)$

$= \alpha\left[\left(\frac{15}{26}\right)\left(\frac{7}{26}\right)\left(\frac{13}{25}\right)\left(\frac{16}{26}\right)\left(\frac{12}{26}\right)\left(\frac{21}{26}\right)\left(\frac{4}{26}\right)\left(\frac{26}{36}\right)\right]$

$= \alpha[0.0003107]$ // — (1)

Illy, $P(A|X) = \alpha\, P(X|A)\, P(A)$

$= \alpha\left[\left(\frac{41}{64}\right)\left(\frac{16}{64}\right)\left(\frac{32}{64}\right)\left(\frac{22}{64}\right)\left(\frac{31}{64}\right)\left(\frac{37}{64}\right)\left(\frac{17}{64}\right)\left(\frac{64}{86}\right)\right]$

$= \alpha[0.0024297]$ // — (2)



(1) + (2) = 1 $\Rightarrow \alpha = 364.9102$ //

HENCE, $P(S|X) = \alpha(0.0003107) = 0.1134$ //

**iii. What is the probability that a person with these attributes will remain in hospital and that s/he will be discharged? What would the Na¨ıve Bayes classifier predict for this person?**

**ANSWER:**

- From the above calculations, we can see that the required probability is given by - P(X | A)
- P(X | A) = 0.8865

**iv. What is the accuracy of the Na¨ıve Bayes classifier? Explain the results in the confusion matrix.**

**ANSWER:**

**Accuracy :**

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          60               69.7674 %
Incorrectly Classified Instances        26               30.2326 %
```

- The accuracy of the Classifier is (74.4186)

**Results in the confusion matrix :**

```
=== Confusion Matrix ===

  a  b   <-- classified as
 58  4 |  a = A
 22  2 |  b = S
```

- **Explanation:**
  - The number of Instances of A which were correctly identified are 58
  - The number of Instances of A which were incorrectly identified is 4

    Similarly,

  - The number of Instances of B which were correctly identified are 22
  - The number of Instances of B which were incorrectly identified are 2

**(c) k-NN (6 + 2 = 8 marks)**

**i. Find three instances in the dataset that are similar to the above patient, and use the Jaccard coefficient to calculate (by hand) the predicted outcome for this patient. Show your calculations.**

Given :

$$X = \begin{cases} \text{L-CORE} = \text{mid} \\ \text{L-SURF} = \text{low} \\ \text{L-O2} = \text{good} \\ \text{L-BP} = \text{high} \\ \text{SURF-STBL} = \text{stable} \\ \text{CORE-STBL} = \text{stable} \\ \text{BP-STBL} = \text{mod-stable} \end{cases}$$

Now,

Solution: 3 similar instances, similar to the above :

$$X1 = \begin{cases} \text{mid} \\ \text{low} \\ \text{good} \\ \text{high} \\ \text{stable} \\ \text{unstable} \\ \text{mod-stable} \\ A \end{cases} \quad X2 = \begin{cases} \text{mid} \\ \text{low} \\ \text{good} \\ \text{high} \\ \text{unstable} \\ \text{unstable} \\ \text{stable} \\ S \end{cases} \quad X3 = \begin{cases} \text{mid} \\ \text{low} \\ \text{good} \\ \text{high} \\ \text{unstable} \\ \text{stable} \\ \text{stable} \\ A \end{cases}$$

Jaccard coefficient $= J(A,B) = \frac{n(A \cap B)}{n(A \cup B)}$

$$\Rightarrow J(X, X1) = \frac{n(X \cap X1)}{n(X \cup X1)} = \frac{6}{6+3} = \frac{6}{9} = 0.66$$

$$J(X, X2) = \frac{n(X \cap X2)}{n(X \cup X2)} = \frac{4}{11} = 0.3636$$

$$J(X, X3) = \frac{n(X \cap X3)}{n(X \cup X3)} = \frac{5}{10} = 0.5\%$$



**ii. What is the accuracy of the k-NN classifier for different values of k? Explain the results in the confusion matrix.**

**Accuracy :**

**For KNN = 1**

```
Correctly Classified Instances          53               61.6279 %
Incorrectly Classified Instances        33               38.3721 %
```

**For KNN = 2**

```
Correctly Classified Instances          61               70.9302 %
Incorrectly Classified Instances        25               29.0698 %
```

**For KNN = 3**

```
Correctly Classified Instances          61               70.9302 %
Incorrectly Classified Instances        25               29.0698 %
```

**Clearly,**
As the value of k increases, the accuracy of the k-NN Classifier increases.
Also, the accuracy saturates at 70.9302%

**Results in the confusion matrix :**

```
=== Confusion Matrix ===

  a  b   <-- classified as
 61  1 |  a = A
 24  0 |  b = S
```

- **Explanation:**
  - The number of Instances of A which were correctly identified are 61
  - The number of Instances of A which were incorrectly identified is 1

  Similarly,

  - The number of Instances of B which were correctly identified are 0
  - The number of Instances of B which were incorrectly identified are 24

**3)**
**(5.5 + 2.5 = 8 marks)**
**Draw a table to compare the performance of J48, Naive Bayes and IBk using the summary measures produced by WEKA. Which algorithm does better? Explain in terms of WEKA's summary measures. Can you speculate why?**

**Comparison Table using the Summary measure :**

| | **Correctly Classified Instances** | **Incorrectly Classified Instances** | **Mean Absolute Error** | **Root Mean Squared error** | **Relative absolute error** | **Root relative squared error** |
|---|---|---|---|---|---|---|
| J48 | 69.7674 % | 30.2326 % | 0.3774 | 0.4957 | 93.1217% | 110.3555% |
| Naive Bayes | 69.7674 % | 30.2326 % | 0.415 | 0.4732 | 102.4041% | 105.3522% |
| IBk | 70.9302 % | 29.0698 % | 0.4242 | 0.5184 | 104.671% | 115.4187% |

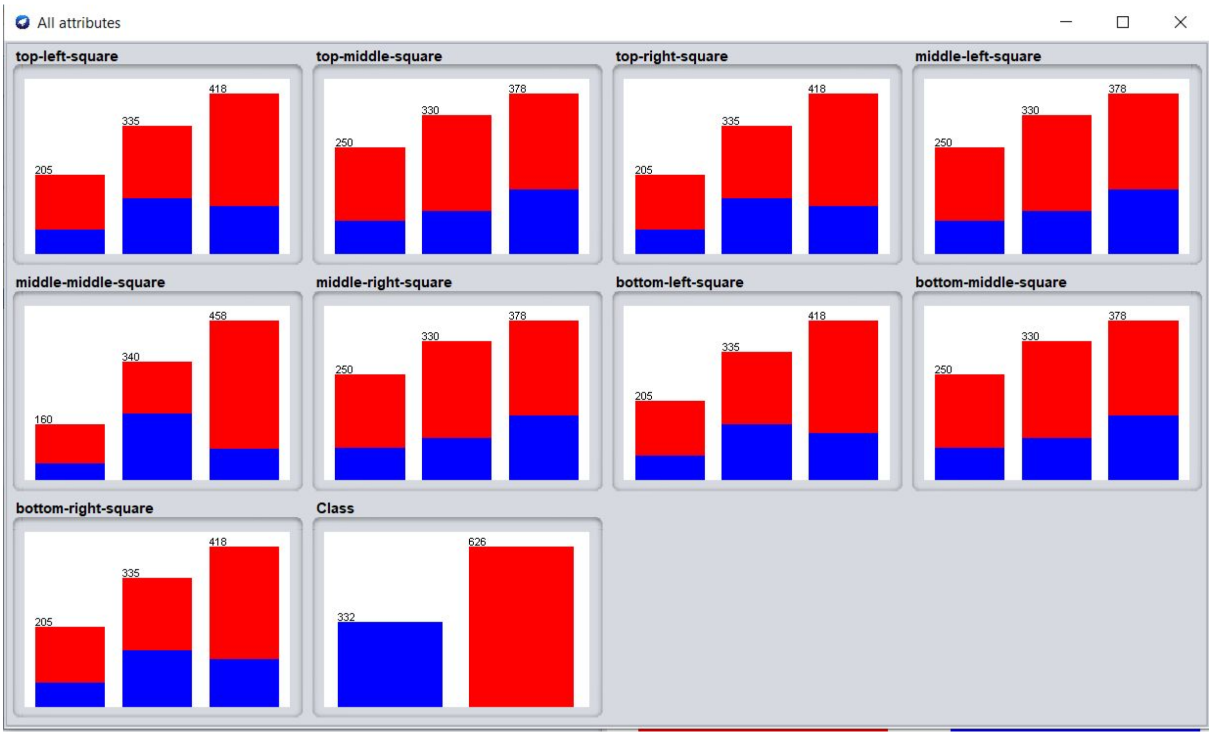
**Explanation:**

- Here, we can see that the **Classification Accuracy** of all the three algorithms are similar and nearby.
- Only the **Classification Accuracy** of IBk is slightly higher than the other two.
- Thus, we can say that Ibk does better in terms of **Classification Accuracy.**

## Question 5: Classification, Decision Trees, Na¨ıve Bayes, k-NN, Weka (52 marks)

Consider the dataset tic-tac-toe.arff available on moodle. Each example in this dataset represents a different game of tic-tac-toe (http://en.wikipedia.org/wiki/Tic-tac-toe), where the player writing crosses ("x") has the first move. Only those games that don't end in a draw are included, with the positive class representing the case where the first player wins and the negative class the case where the first player loses. The features encode the status of the game at the end, so each square contains a cross "x", a nought "o" or a blank "b".

**1. Before you run the classifiers, use the WEKA visualization tool to analyze the data. (2+2 = 4 marks)**
**(a) Which attributes seem to be the most predictive of winning or losing? (hint: if you were the "x" player, where would you put your first cross and why?)**
**ANSWER :**




**Most Important Attribute :**
If I was player x, I would start with either of the following:
- Top Middle Square
- Bottom Middle Square
- Middle Left Square

Because these squares have higher chances of x winning.

**(b) What can you infer about the advantage (or otherwise) of being the first player?**
**ANSWER :**

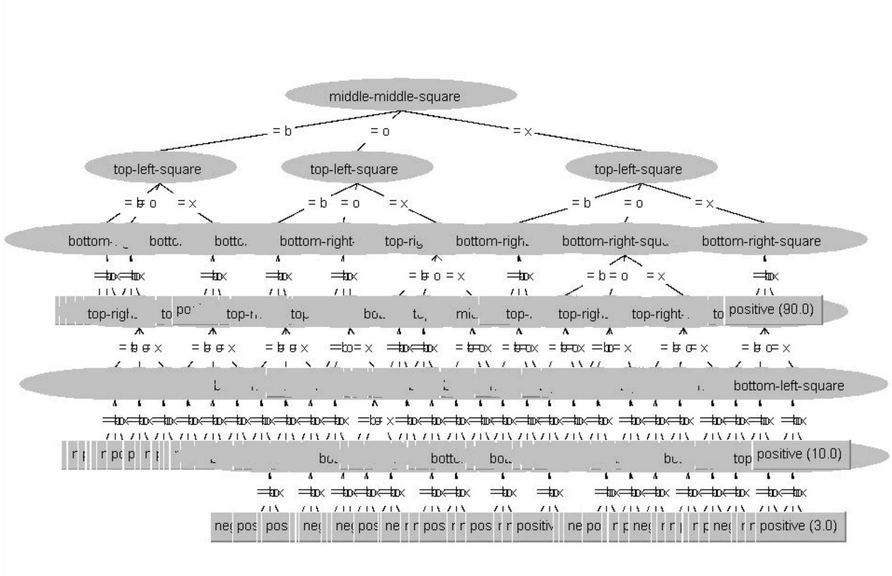
- We can see that the advantage of being the first player is that most of the time the first player wins.

**2. Run J48 (=C4.5, decision tree), Na¨ıve Bayes and IBk (=k-NN) to learn a model that predicts whether the "x" player will win. Perform 10-fold cross validation, and analyze the results obtained by these algorithms as follows.**

**Note: When using IBk, click on the "Choose" bar to set the value of k (default is 1). Consider different values of k**

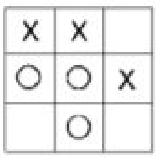
**<u>(a) J48 (=C4.5) (2 + 3 + 14 + 3 = 22 marks)</u>**

**i. Examine the decision tree and indicate the main variables.**
**ANSWER :**




**Main Variables:**
- Middle-middle square - as this has the most effect in deciding the outcome.
- Also from the combined visualisation we can see that, in that box - x has won the most number of times (458 times)

**ii. Trace the decision tree for the following game. What would it predict?**



- Prediction would be **Negative**

**iii. What is the first split in the decision tree? Calculate (by hand) the Information Gain obtained from the first split in the tree. Show your calculations.**

Information gain obtained:

Initial Entropy = ?

Solution:

$H(S)$ = Entropy of class

$$= \left(\frac{-P}{P+N}\right)\log_2\left(\frac{P}{P+n}\right) - \left(\frac{N}{P+N}\right)\log_2\left(\frac{N}{P+N}\right)$$

$$= \sum p \log_2(p)$$

$$\Rightarrow H(S) = \frac{-626}{958}\log_2\left(\frac{626}{958}\right) - \left(\frac{332}{958}\right)\log_2\left(\frac{332}{958}\right)$$

$$= (-0.65)(-0.63) - (0.346)(-1.528)$$

$$H(S) = 0.93 \quad \text{—— (1)}$$

Entropy of Middle-Middle square:

1) $H(b) = \frac{-112}{160}\log_2\left(\frac{112}{160}\right) - \frac{48}{160}\log_2\left(\frac{48}{160}\right)$

↑ (BLANK)

$= 0.87$

2) $H(X) = -\frac{366}{458}\log_2\left(\frac{366}{458}\right) - \frac{92}{458}\log_2\left(\frac{92}{458}\right)$

↑ CROSS

$= 0.732$

3) $H(O) = \frac{-148}{340}\log_2\left(\frac{148}{340}\right) - \frac{192}{340}\log_2\left(\frac{192}{340}\right)$

$= 0.98$

Required Entropy:

$H(Blank) = H(b) = 0.87$

$H(0) = 0.98$

$H(X) = 0.72$

Entropy

$$= \left(\frac{112+49}{958}\right)(0.87)\left(\frac{148+192}{958}\right)(0.98)$$

$$+ \left(\frac{366+92}{958}\right)(0.72)$$

$$= 0.843 \text{ —— } (2)$$

Now,

Information Gain

$= H(S) - \text{Entropy}$

$= 0.087$

Hence, Information GAIN $= 0.087$

**iv. What is the accuracy of the decision tree? Explain the results in the confusion matrix.**

**Accuracy : 85.3862%**

```
=== Summary ===

Correctly Classified Instances         818               85.3862 %
Incorrectly Classified Instances       140               14.6138 %
```

**Results in the confusion matrix :**

```
=== Confusion Matrix ===

   a   b   <-- classified as
 269  63 |   a = negative
  77 549 |   b = positive
```

- **Explanation:**
    - The number of Instances of A which were correctly identified are 269
    - The number of Instances of A which were incorrectly identified is 63

    Similarly,

    - The number of Instances of B which were correctly identified are 549
    - The number of Instances of B which were incorrectly identified are 24

**<u>(b) Na¨ıve Bayes (6 + 2 + 2 = 10 marks)</u>**

**i. Calculate (by hand) the predicted probability of a win for the following game. Show your calculations.**

b) i)

| | | |
|---|---|---|
| X | X | |
| 0 | 0 | X |
| | 0 | |

PREDICTED PROBABILITY OF A WIN FOR THE FOLLOWING GAME:-

CONDITIONS:-

$X = [$ Top-left-square = x,
Top-middle-square = x,
middle-right-square = x,
middle-middle square = 0,
middle-left-square = 0,
bottom-middle-square = 0,
Top-right-square = b,
Bottom-right-square = b,
Bottom-left-square = b$]$

$$P(X) = \frac{1}{\alpha}$$

$$P(\text{Positive}|X) = \frac{P(X|\text{Positive})\, P(\text{positive})}{P(X)}$$

$$P(\text{Negative}|X) = \frac{P(X|\text{Negative})\, P(\text{Negative})}{P(X)}$$

$$P(\text{Positive}|X) = \left(\frac{296}{629}\right)\left(\frac{226}{629}\right)\left(\frac{143}{629}\right)\left(\frac{230}{629}\right)\left(\frac{149}{629}\right)\left(\frac{226}{629}\right)\left(\frac{143}{629}\right)\left(\frac{230}{629}\right)\left(\frac{143}{629}\right)\left(\frac{629}{964}\right)$$

$$= 0.0001470\alpha \quad \text{—— (1)}$$

$$P(\text{Negative}|X) = \left(\frac{124}{335}\right)\left(\frac{154}{335}\right)\left(\frac{64}{335}\right)\left(\frac{102}{335}\right)\left(\frac{193}{335}\right)\left(\frac{154}{335}\right)\left(\frac{64}{335}\right)\left(\frac{102}{335}\right)\left(\frac{64}{335}\right)\left(\frac{335}{964}\right) = 0.000101 \quad \text{—— (2)}$$

① + ② = 1

$$0.000147\alpha + 0.000101\alpha = 1$$

$$\alpha = \frac{1}{(2.48 \times 10^{-4})}$$

$$= 4032 //.$$

So, $P(\text{positive} \mid X)$

$$= 0.000147\alpha$$

$$= 0.5927 //.$$

$P(\text{Negative} \mid X)$

$$= 0.000101\alpha$$

$$= 0.4072 //.$$



**ii. What is the probability that a player with this configuration will win? What would the Na¨ıve Bayes classifier predict for this game?**

Naive Bayes Classifier's Prediction is around - 0.594

We can see that it is closer to the theoretical calculated value.

**iii. What is the accuracy of the Na¨ıve Bayes classifier? Explain the results in the confusion matrix.**

**ANSWER :**

Accuracy : 69.6242%

```
=== Summary ===

Correctly Classified Instances         667               69.6242 %
Incorrectly Classified Instances       291               30.3758 %
```

**Results in the confusion matrix:**

```
=== Confusion Matrix ===

   a   b   <-- classified as
 142 190 |   a = negative
 101 525 |   b = positive
```

- **Explanation:**
    - The number of Instances of A which were correctly identified are 142
    - The number of Instances of A which were incorrectly identified is 190

    Similarly,

    - The number of Instances of B which were correctly identified are 525
    - The number of Instances of B which were incorrectly identified are 101

**(c) k-NN (6 + 2 = 8 marks)**

**i. Find three instances in the dataset that are similar to the following game, and use the Jaccard coefficient to calculate (by hand) the predicted outcome for this game. Show your calculations.**

c) Given:

i) $X = \{x, x, b, o, o, x, b, o, b\}$

Instance1: $X1 = \{x, x, b, o, o, x, x, o\}$

Instance2: $X2 = \{x, x, b, o, o, o, x, o, x\}$

Instance 3: $X3 = \{x, x, b, o, o, o, b, x, b\}$

Solution:

Jaccard coefficient

$$= J(A,B) = \frac{n(A \cap B)}{n(A \cup B)}$$

$$\Rightarrow J(X, X1) = \frac{n(X \cap X1)}{n(X \cup X1)}$$

$$= \frac{5}{13} = 0.8$$

$$J(X, X2) = \frac{n(X \cap X2)}{n(X \cup X2)}$$

$$= \frac{6}{11}$$

$$= 0.54$$

$$J(X, X3) = \frac{n(X \cap X3)}{n(X \cup X3)}$$

$$= \frac{7}{11}$$

$$= 0.63$$



**ii. What is the accuracy of the k-NN classifier? Explain the results in the confusion matrix.**
**ANSWER :**

**Accuracy : 98.9562%**
**Results in the confusion matrix:**

```
=== Confusion Matrix ===

   a   b   <-- classified as
 323   9 |   a = negative
   1 625 |   b = positive
```

- **Explanation:**
    - The number of Instances of A which were correctly classified are 323

- The number of Instances of A which were incorrectly classified is 9

  Similarly,

- The number of Instances of B which were correctly classified are 625
- The number of Instances of B which were incorrectly identified are 1

**3. (5.5 + 2.5 = 8 marks)**
**Draw a table to compare the performance of J48, Na¨ıve Bayes and IBk using the summary measures produced by WEKA. Which algorithm does better? Explain in terms of WEKA's summary measures. Can you speculate why?**

| | **Correctly Classified Instances** | **Incorrectly Classified Instances** | **Mean Absolute Error** | **Root Mean Squared error** | **Relative absolute error** | **Root relative squared error** |
|---|---|---|---|---|---|---|
| J48 | 85.3862 % | 14.6138 % | 0.1485 | 0.343 | 32.7887% | 72.077% |
| Naive Bayes | 69.6242 % | 30.3758 % | 0.3702 | 0.4319 | 81.7182% | 90.7504% |
| IBk | 98.9562 % | 1.0438 % | 0.1909 | 0.2315 | 42.1333% | 48.6378% |

- Thus, we can see that IBk has the best **Classification Accuracy.**
- Naive Bayes has the least correct **Classification Accuracy** followed by J48 and finally topped by IBk.
- In this case, Naive Bayes has the **highest amount** of Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error and Root Relative Squared Error.

**Question 6: Regression (1 + 2 + 5 = 8 marks)**

Consider the dataset abs.arff available on moodle. This dataset contains continuous-valued economic attributes of a country, with the target variable being the unemployment rate. Additional documentation regarding these attributes appears in the arff file.

1. Perform a linear regression (under functions in WEKA) to learn a linear model of the unemployment rate as a function of the other variables. You can use the default parameters given in WEKA. What is the resultant regression function?

**ANSWER :**

The Regression Model is given by the following formula :

```
Linear Regression Model

Unemployment-Rate =

      -0.0014 * All-Ords-Index +
      -0.2452 * Housing-Loan-Interest-Rate +
      13.7286
```

2. Using the resultant regression function, calculate by hand the Absolute Error for the year 1986.

**ANSWER :**

For the year 1986, we have :

| | | | | | |
|---|---|---|---|---|---|
| 1986 | 204.4 | 1779.1 | 15.5 | 77547 | 8.4 |

Absolute Error

= | Measured Value - Given Value |

Here,

Given Value = 8.4

Measured Value

= -0.0014 * (All-Ords-Index) - 0.2452 * (Housing-Loan-Interest-Rate) + 13.7286

= -0.0014 * (1779.1) - 0.2452 * (15.5) + 13.7286

= 7.43726

Now, Absolute Error

= | Measured Value - Given Value |

= | 7.43726 - 8.4 |

= 0.96248

3. Calculate the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) obtained by the regression function (you can use the excel spreadsheet provided on moodle). How is MAE different from RMSE? (do these functions emphasize different aspects of performance?)

**ANSWER :**

**Mean Absolute Error (MAE)**

= sigma(| y’ - y |)/n)

= (21.54996/21)

= 1.0261

**Root Mean Square Error (RMSE)**

= [ sigma| y’ - y |^2)/n ]^(0.5)

= (34.47705959/21)^(0.5)

= 1.2813

The above values were calculated from Excel.

- I removed all rows with missing values and calculated the above for the years between 1996 to 2006.

| x1 | x2 | | y | y' | y - y' | | | \| y - y' \| | \|y - y'\|^2 |
|---|---|---|---|---|---|---|---|---|---|
| 1779.1 | 15.5 | 77547 | 8.4 | 7.43726 | 0.96274 | | | 0.96274 | 0.926868308 |
| 1585.3 | 13.5 | 84101 | 7.7 | 8.19898 | -0.49898 | | | 0.49898 | 0.24898104 |
| 1527.7 | 17 | 84981 | 6.5 | 7.42142 | -0.92142 | | | 0.92142 | 0.849014816 |
| 1508.8 | 16.5 | 88950 | 6 | 7.57048 | -1.57048 | | | 1.57048 | 2.46640743 |
| 1504.9 | 13 | 98970 | 8.5 | 8.43414 | 0.06586 | | | 0.06586 | 0.00433754 |
| 1652.7 | 10.5 | 108328 | 11.1 | 8.84022 | 2.25978 | | | 2.25978 | 5.106605648 |
| 1722.6 | 9.5 | 116355 | 11.8 | 8.98756 | 2.81244 | | | 2.81244 | 7.909818754 |
| 2040.2 | 8.75 | 127362 | 11.3 | 8.72682 | 2.57318 | | | 2.57318 | 6.621255312 |
| 2000.8 | 10.5 | 132979 | 9.3 | 8.35288 | 0.94712 | | | 0.94712 | 0.897036294 |
| 2231.7 | 9.75 | 146225 | 8.8 | 8.21352 | 0.58648 | | | 0.58648 | 0.34395879 |
| 2662.7 | 7.2 | 162144 | 8.9 | 8.23538 | 0.66462 | | | 0.66462 | 0.441719744 |
| 2608.2 | 6.7 | 169720 | 8.6 | 8.43428 | 0.16572 | | | 0.16572 | 0.027463118 |
| 2963 | 6.5 | 172977 | 7.8 | 7.9866 | -0.1866 | | | 0.1866 | 0.03481956 |
| 3115.9 | 7.8 | 189827 | 6.8 | 7.45378 | -0.65378 | | | 0.65378 | 0.427428288 |
| 3352.4 | 6.8 | 205369 | 6.6 | 7.36788 | -0.76788 | | | 0.76788 | 0.589639694 |
| 3241.5 | 6.55 | 203964 | 7 | 7.58444 | -0.58444 | | | 0.58444 | 0.341570114 |
| 3032 | 6.55 | 204334 | 6.4 | 7.87774 | -1.47774 | | | 1.47774 | 2.183715508 |
| 3499.8 | 7.05 | 206761 | 5.8 | 7.10022 | -1.30022 | | | 1.30022 | 1.690572048 |
| 4197.5 | 7.3 | 213985 | 5.2 | 6.06214 | -0.86214 | | | 0.86214 | 0.74328538 |
| 4933.5 | 7.55 | 219678 | 4.9 | 4.97044 | -0.07044 | | | 0.07044 | 0.004961794 |
| 6337.6 | 8.05 | 228442 | 4.5 | 2.8821 | 1.6179 | | | 1.6179 | 2.61760041 |
| | | | | | | | | 21.54996 | 34.47705959 |

**How is MAE different from RMSE :**

- For both MAE and RMSE, the lower their value, the better is the measure.
- In RMSE,
  - We calculate the squares of the errors
  - And, then calculate their average
  - Hence, RMSE becomes a useful measure when we are dealing with large, undesirable errors.
- However, MAE is simple and interpretation of MAE is easier.