AI course - Lab 3

Fredrik Fredriksson

December 2019

1 Part 1

1.1 Task A

Attempt 1:

- Log Likelihood: -4700,03365589168
- BIC: 9590,04864149962

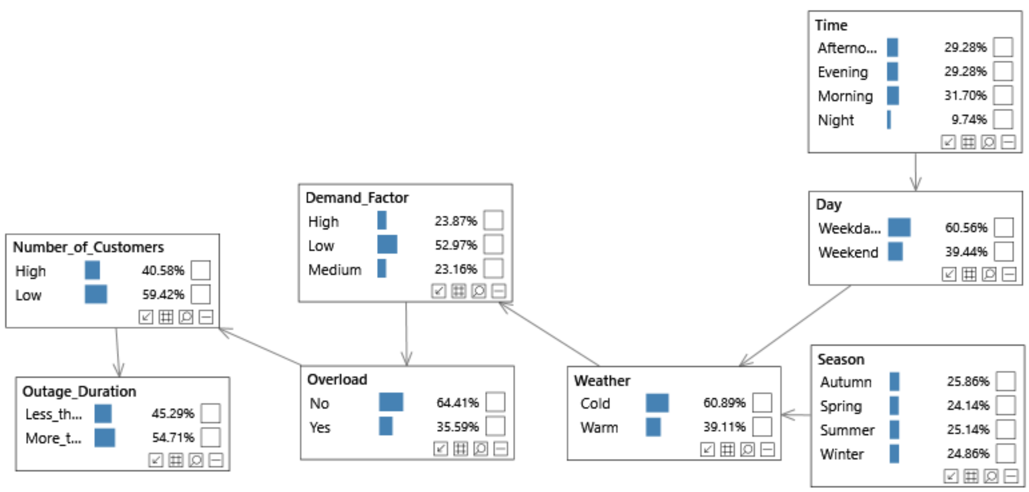

Figure 1: Bayers Network

1.2 Task B

1.2.1 Log Likelihood

- the log likelihood of the data, given the candidate network. Log likelihood estimation is a method that determines the values for a models parameters. The natural logarithm of the expression is preferred since its a monotonically increasing function. This results in when x-axis increases, the value on the y-axis also increases. This will ensure that the maximum value of the log probability occurs at the same point as the original probability function. The parameter values are decided so they will maximise the likelihood of a process that is described by the model produced from the data that were actually observed. In other words its purpose is to figure out which type of model that describes the data points best. The model with the highest log likelihood is the most preferred model.

1.2.2 BIC

- Bayesian Information Criterion Bayesian information criterion is a criterion for model selection among a finite set of models. The model with the lowest BIC is preferred. Since it is partly based on the Log likelihood it is possible to implement more parameters, this may however cause an overfit. Therefor the BIC try to avoid this by introducing an penalty term for the number of parameters in the model.

To sum it up it is possible to say that Log likelihood describe how well the model fits and BIC how complex the model is.

1.3 Task C

1.3.1 Attempt 2

- Log Likelihood: -4833,05373490735
- BIC: 9797,12907651557

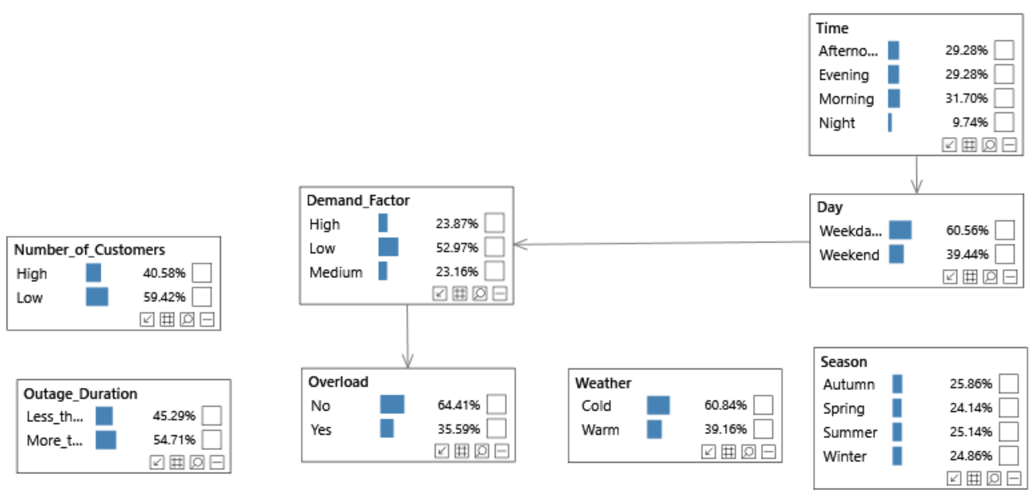

Figure 2: Bayers Network

1.3.2 Attempt 3

- Log Likelihood: -4472,1302388737
- BIC: 9284,91665516967

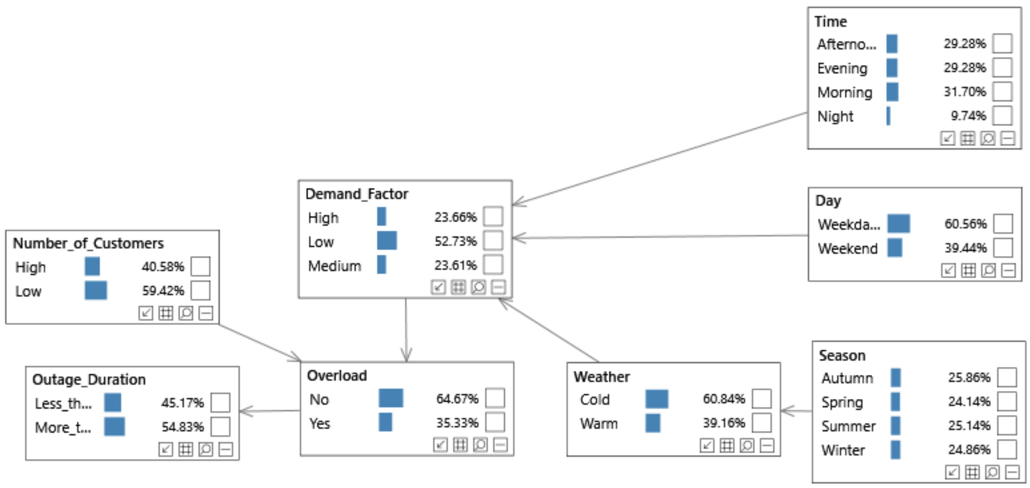

Figure 3: Bayers Network

The the models are almost equally bad. Either they have a high complexity, correct model or both. By observing the models attempt 3 in Task B is most likely the best one. This is since it have a combination of high Log likelihood and low BIC compared to the others were the likelihood is lower and BIC higher. This does however not mean that the model is good at all.

1.4 Task D

- Log Likelihood: -4468,78032075825
- BIC: 9186,50169424815

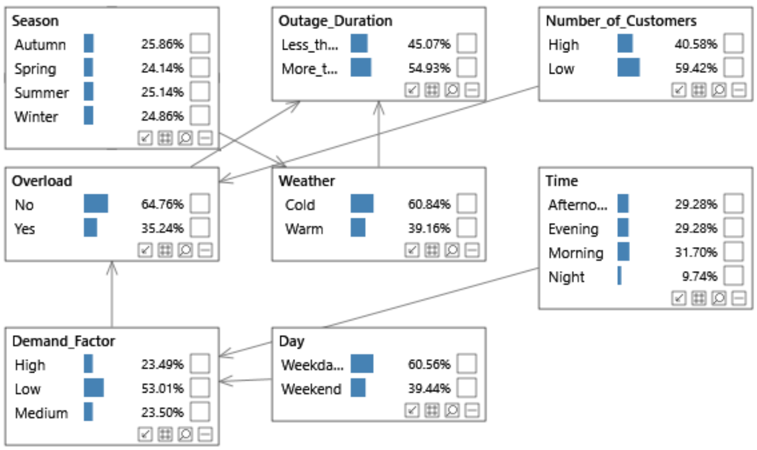

Figure 4: Bayers Network

This model were a little better than the previous but the difference between attempt 3 and the one in Task D isn't very big. This model does however have lower BIC and just a little lower Log likelihood which mean that is performs overall better than the previous ones.

2 Part 2

2.1 Task A

2.1.1 Attempt 1

- Log Likelihood: -14747,1692447784
- BIC: 58537,5479098341

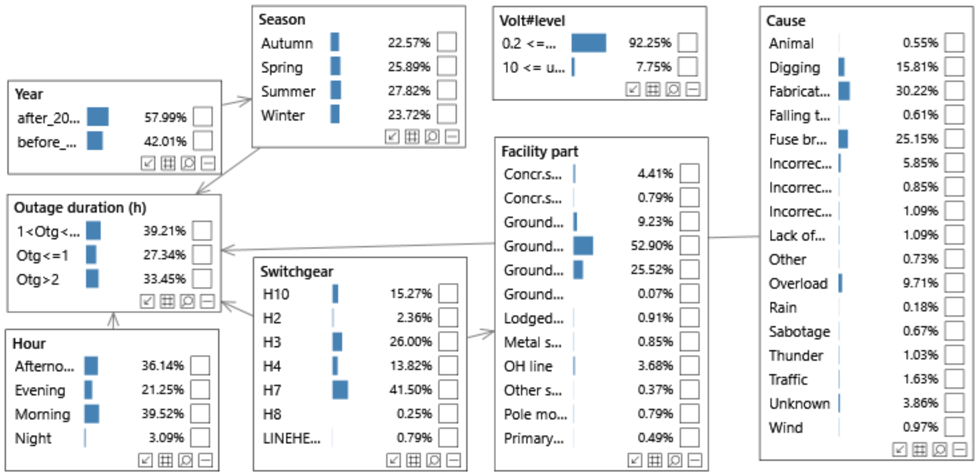

Figure 5: Bayers Network

2.1.2 Attempt 2

- Log Likelihood: -14487,7577991444
- BIC: 368035,34175538

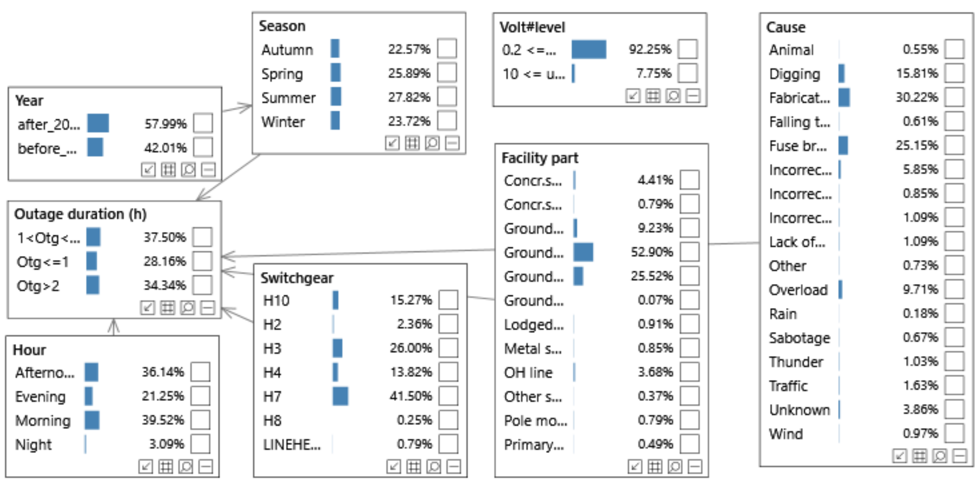

Figure 6: Bayers Network

2.1.3 Attempt 3

- Log Likelihood: -14571,843681523

- BIC: 86392,4638696312

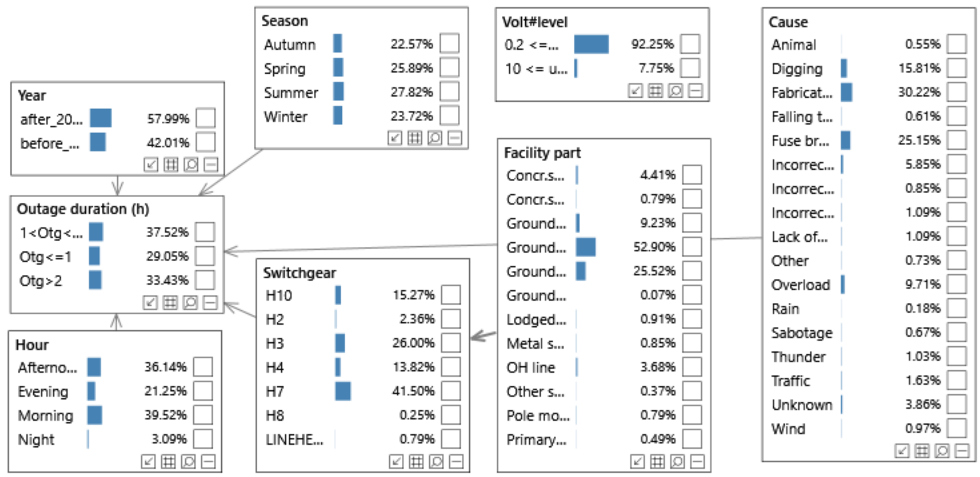

Figure 7: Bayers Network

2.1.4 Attempt 4

- Log Likelihood: -14090,5074716421
- BIC: 69959,3529455002

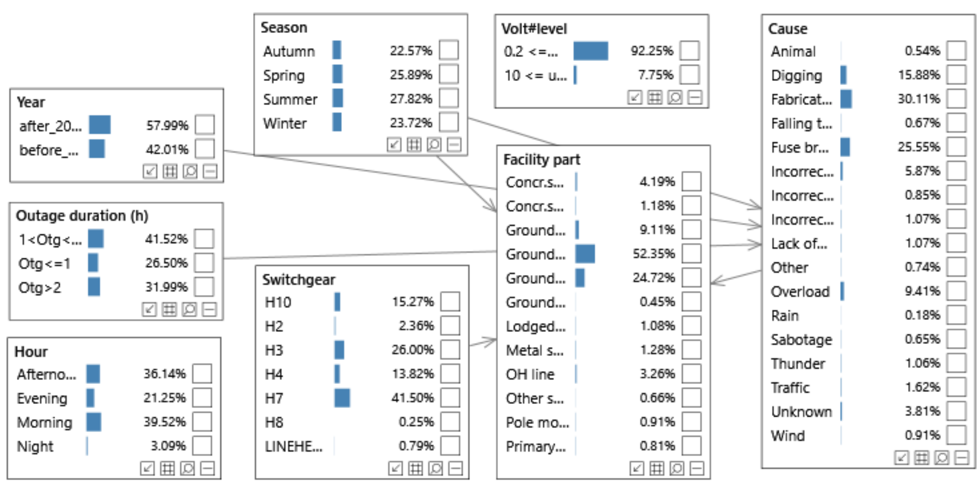

Figure 8: Bayers Network

Attempt 2, 3 did not perform very will. With high BIC compared to the others and average Log likelihood made them unreasonable to use. However attempt 1 performed best in BIC with a little lower Log likelihood compared to attempt 4. But since we cant confirm that is used all available data it can not be decided to be a good model. In disguise it may be over-fitting without our knowledge since it could have excluded important data.

2.2 Task B

1. p(Cause=Animal | Season=Autumn)=0.81%
2. p(Season=Autumn | Cause=Animal)=33.64%
3. p(Season=Summer | Cause=Thunder)=75.33%
4. p(Outage duration=Otg ≤ 1 | Facility part = Ground cable pillar)=29.85%
5. p(Facility part=Ground cable pillar | Switchgear=H7, Cause=Fuse break)=81.79%
6. p(Facility part=Ground feeder cable in ground|Cause=(Digging, Fabrication fault), Switchgear=H7, Season=Summer)=49.61%
7. p(Facility part=Ground feeder cable in ground | Cause=$\neg$(Digging, Fabrication fault), Switchgear=H7, Season=Summer)=27.08%
8. p(Cause=Digging | Facility part = OH line, Switchgear=H7)=0.0%
9. p(Facility part = Ground cable pillar | Outrage duration=Otg ¿ 2=45.21%
10. p(Cause=Unknown | Y ear=bef ore2011)=5.81%
11. p(Cause=Unknown | Y ear=af ter2011) =2.36%

2.3 Task C

1. p(Switchgear=H10 | Cause=Fuse break)=25.55%
2. p(Cause=Wind | Season=Winter)=1.83%
3. p(Facility=OH Line | Season=Winter, Cause=Traffic)=8.33%
4. p(Facility=OH Line | Cause=Rain, Season=Autumn)=8.33%
5. p(Facility=Primary substation | Cause=Rain, Season=Autumn)=8.33%
6. p(Outage Duration=Otg¿2 | Cause=Overload)=18.69%
7. p(Outage Duration=Otg¿2 | Cause=Animal)=44.22%
8. p(Outage Duration=Otg¿2 | Cause=Animal, Season=Winter)=55.74%
9. p(Outage Duration=Otg¿2 | Cause=Animal, Season=Spring)=0.22%
10. p(season=Winter | Outage Duration=Otg¡=1, Cause=Digging)=0.02%
11. p(Cause=Thunder | switchgear=H8, Season=Summer)=2.86%

2.4 Task D

This is due to the fact that the model is an causal network. That means that the relationship between the nodes is causal. By modifying the probability of a node the probability dense function will change and cutting the links and replacing them with the modified value.

These values is observed:

1. p(Cause=Winter)=23.72%
2. p(Cause=Summer)=27.82%
3. p(Cause=Summer | Cause=Digging)=29.72%
4. p(Cause=Winter |Cause=Digging)=14.07%

This means that when the child in a model changes the probability that it was caused by a specific parent changes. However since more parameters could be a cause of it we cannot exclude other information.

2.5 Task E

1. p(Facility part=Ground feeder cable in ground | Cause=Digging, Season=Summer, SwitchgearLINEHED EON) = 99.89%
2. p(Cause=Fabrication fault | Year=After_2011, Season=Spring, Outage Duration=Otg > 2)=64.97%

2.6 Task F

High correlation:

1. p(Cause=Digging | Switchgear=H7)=15.88%
2. p(Cause=Thunder | Season=Summer)=2.86%
3. p(Cause=Rain | Season=Spring)=0.48%

Low correlation:

1. p(Cause=Thunder | Season=Winter)=0%
2. p(Cause=Thunder | Season=Autumn)=0%
3. p(Facility part=Ground cable pillar | Cause=Animal)=0.04%