Rui Carapinha **Artificial Intelligence**

248728 **Bayesian Networks**

1.

The dataset is about the weather in Madrid from 1997 to 2015 (https://www.kaggle.com/juliansimon/weather_madrid_lemd_1997_2015.csv). It contains the following variables: Date, Temperature (Max, Min and Mean), Dew Point (Max, Min and Mean), Humidity (Max, Min and Mean), Sea Level Pressure (Max, Min and Mean), Visibility (Max, Min and Mean), Max Gust Speed, Precipitation, Cloud Cover and Wind Direction. For this network I'm going to use the mean values of Temperature, Dew Point, Humidity, Sea Level Pressure, Visibility, Wind Speed and the values of Precipitation and Wind Direction.

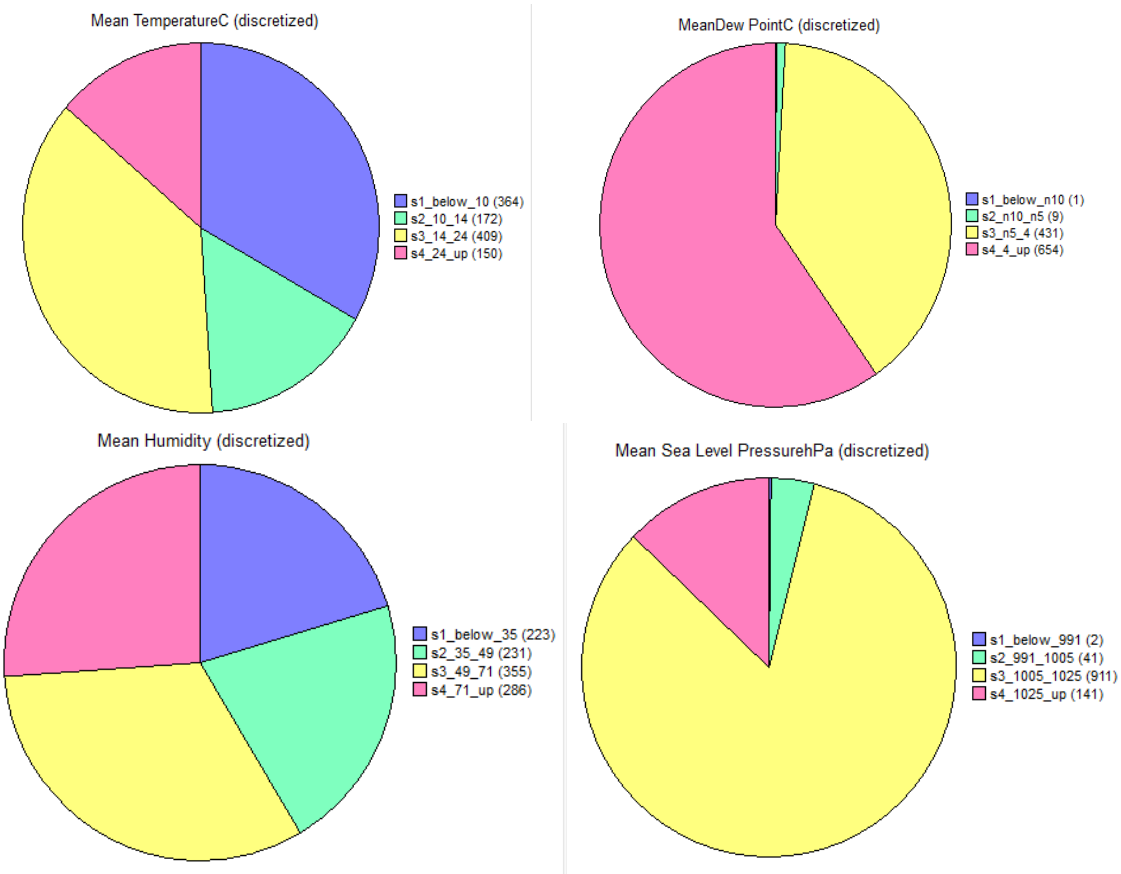
I'm only using the values since 2013 (around 1000 values) because it was very hard to compute all the values (around 6000).

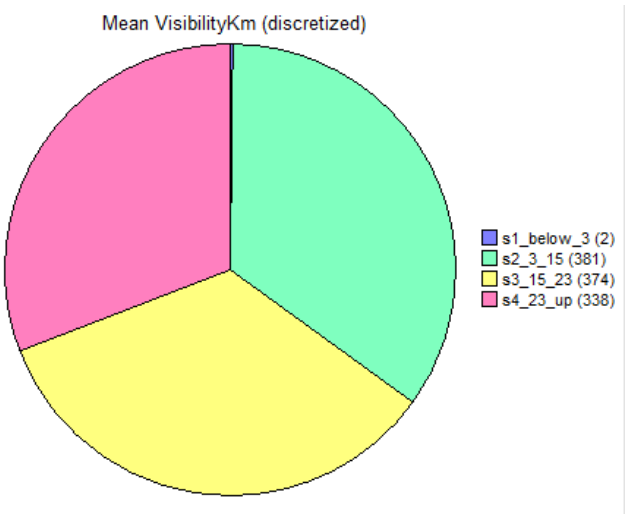
The decision node I included is about if I should bring or not a jacket.

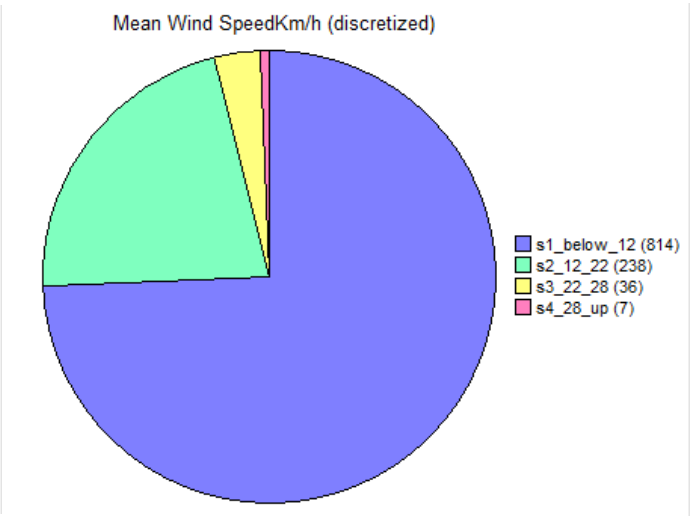
2.

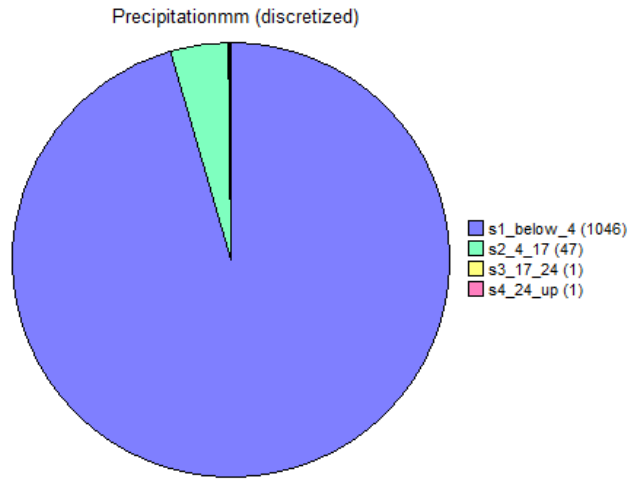
For this part of the assignment I used the software Genie to help me build and verify the belief network model.

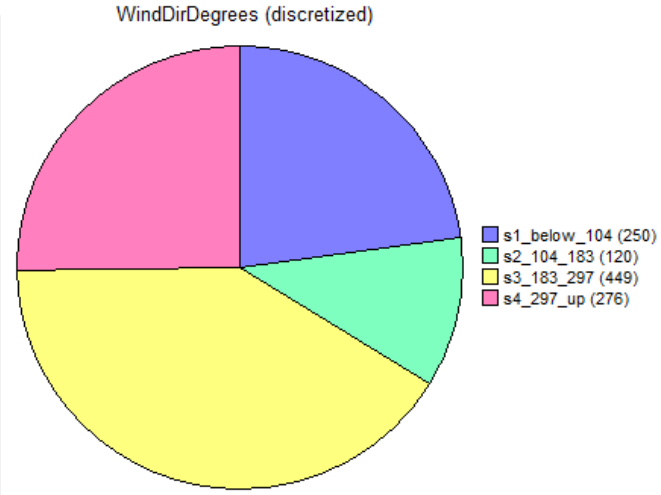
The probability distribution of each variable is the following:











The probabilities obtained for each variable were the following:

a) Mean Dew Point:

	Mean Visibility			
Mean Dew Point	**< 3**	**3 - 15**	**15 - 23**	**>23**
< -10	0.083	0.001	0.001	0.001
-10 - -5	0.083	0.003	0.009	0.015
-5 - 4	0.417	0.372	0.369	0.443
> 4	0.417	0.624	0.622	0.538

b) Mean Temperature:

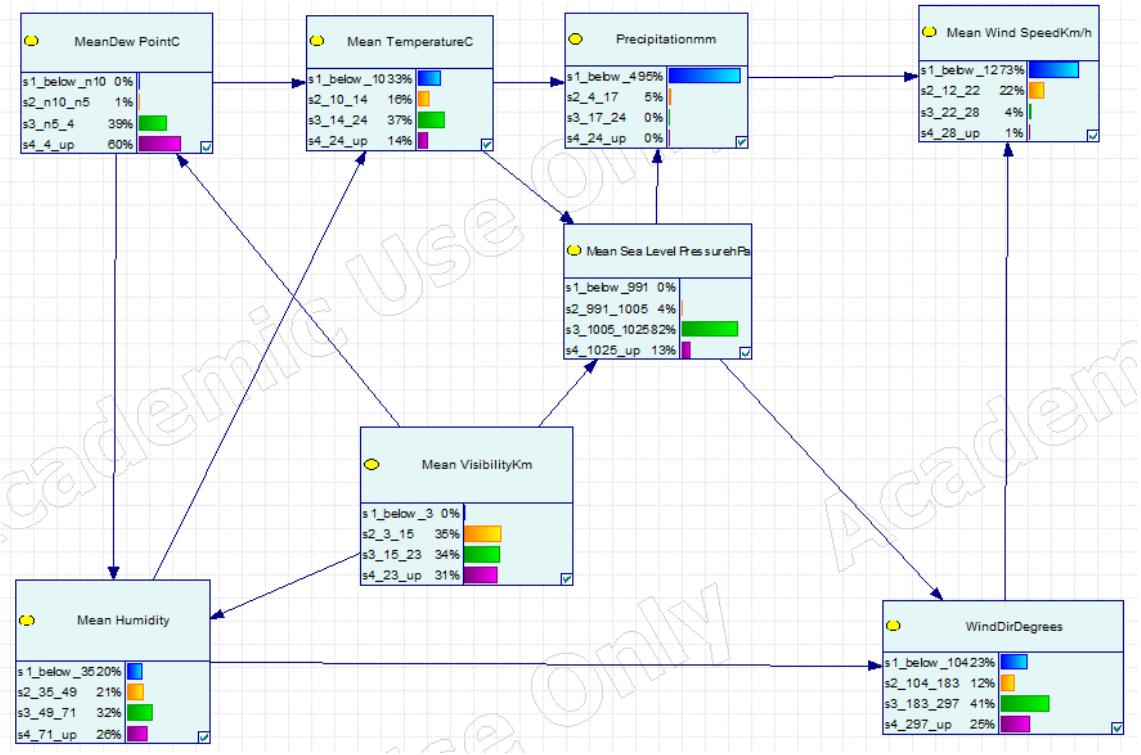
Mean Dew Point		**< -10**				**-10 - -5**			
Mean Humidity		**< 35**	**35-49**	**49-71**	**>71**	**< 35**	**35-49**	**49-71**	**>71**
Mean Temperature	**< 10**	0.625	0.25	0.25	0.25	0.625	0.85	0.85	0.25
	10-14	0.125	0.25	0.25	0.25	0.125	0.05	0.05	0.25
	14-24	0.125	0.25	0.25	0.25	0.125	0.05	0.05	0.25
	> 24	0.125	0.25	0.25	0.25	0.125	0.05	0.05	0.25
Mean Dew Point		**-5 - 4**				**> 4**			
Mean Humidity		**< 35**	**35-49**	**49-71**	**>71**	**< 35**	**35-49**	**49-71**	**>71**
Mean Temperature	**< 10**	0.004	0.187	0.842	0.994	0.002	0.002	0.008	0.391
	10-14	0.038	0.384	0.150	0.002	0.002	0.008	0.274	0.402
	14-24	0.754	0.426	0.007	0.002	0.302	0.837	0.711	0.205
	> 24	0.204	0.004	0.001	0.002	0.695	0.153	0.008	0.001

c) Mean Humidity:

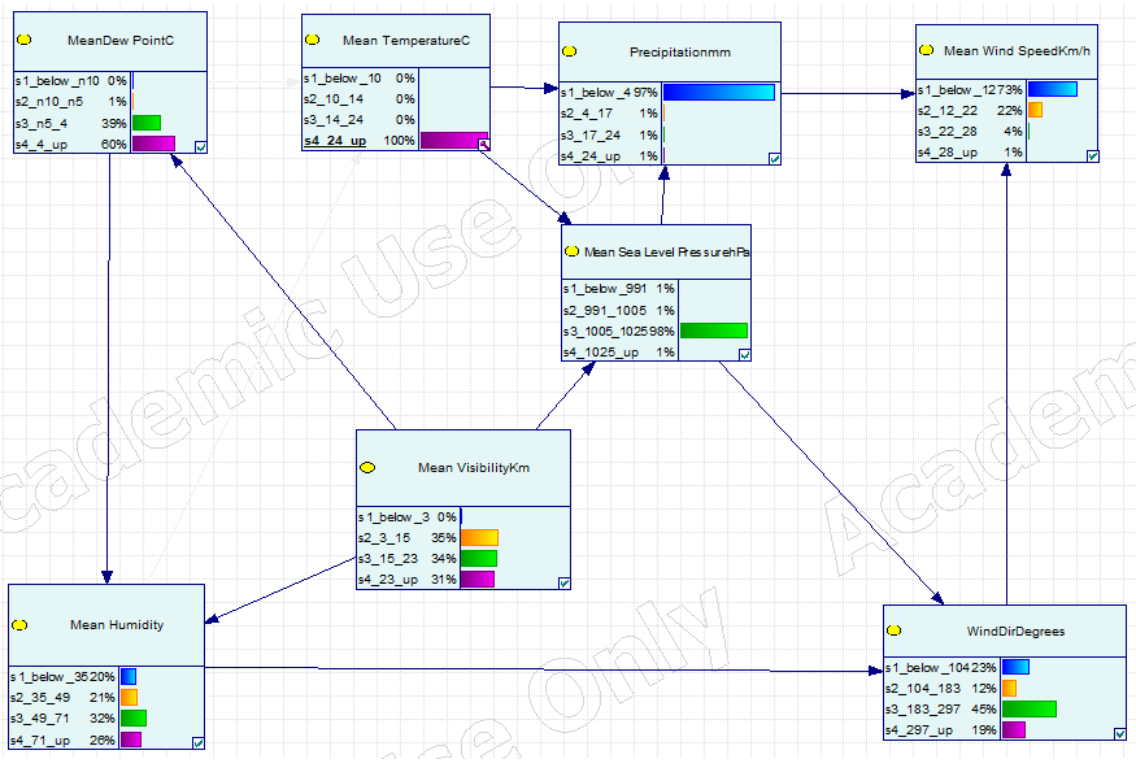
Mean Dew Point		**< -10**				**-10 - -5**			
Mean Visibility		**< 3**	**3-15**	**15-23**	**>23**	**< 3**	**3-15**	**15-23**	**>23**
Mean Humidity	**< 10**	0.25	0.25	0.25	c	0.25	0.125	0.063	0.208
	10-14	0.25	0.25	0.25	0.125	0.25	0.125	0.563	0.375
	14-24	0.25	0.25	0.25	0.125	0.25	0.625	0.313	0.375
	> 24	0.25	0.25	0.25	0.125	0.25	0.125	0.063	0.042
Mean Dew Point		**-5 - 4**				**> 4**			
Mean Visibility		**< 3**	**3-15**	**15-23**	**>23**	**< 3**	**3-15**	**15-23**	**>23**
Mean Humidity	**< 10**	0.125	0.002	0.052	0.346	0.125	0.068	0.206	0.537
	10-14	0.125	0.044	0.167	0.273	0.125	0.122	0.364	0.236
	14-24	0.125	0.442	0.585	0.286	0.125	0.210	0.330	0.204
	> 24	0.625	0.512	0.196	0.094	0.625	0.599	0.099	0.023

3.

Probability distribution before querying the network:

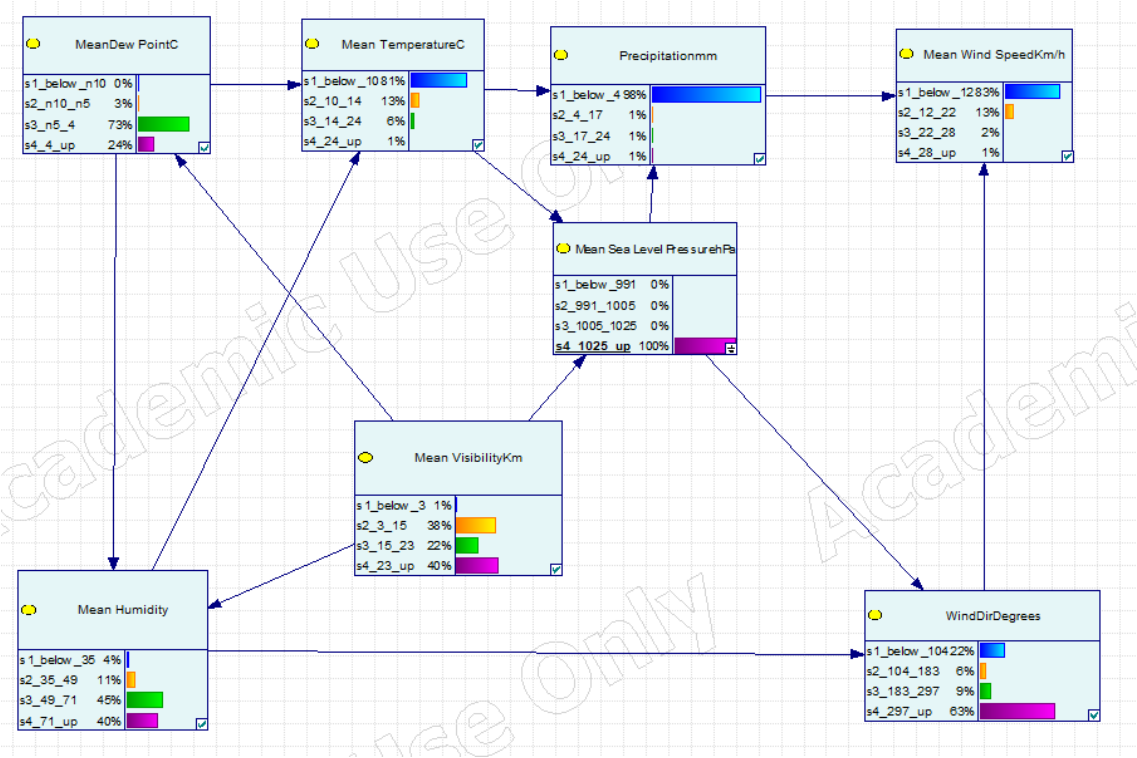


a. Query with Temperature > 24:



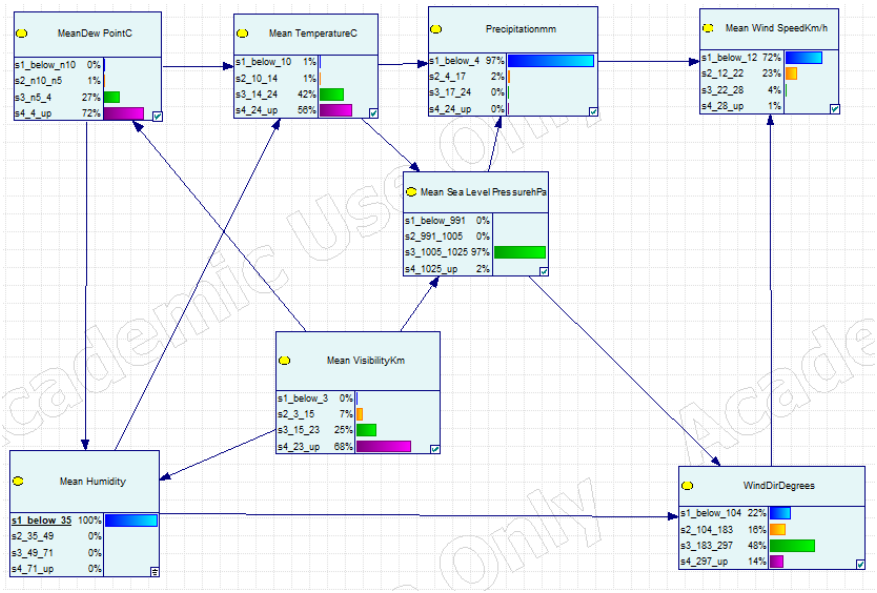
In this query we don't see a lot of differences. The biggest difference we see is that the Mean Sea Level Pressure is 98% probable to be between 1005 and 1025.

b. Query with Sea Level Pressure > 1025:



Here we can see that when Mean Sea Level Pressure is bigger than 1025, it changes the wind direction (63% of being bigger than 297 degrees). We can see that affects a lot the temperature of Madrid. When the Mean Sea Level Pressure is bigger than 1025 we almost always (around 81% of probability) of having temperatures lower than 10ºC.

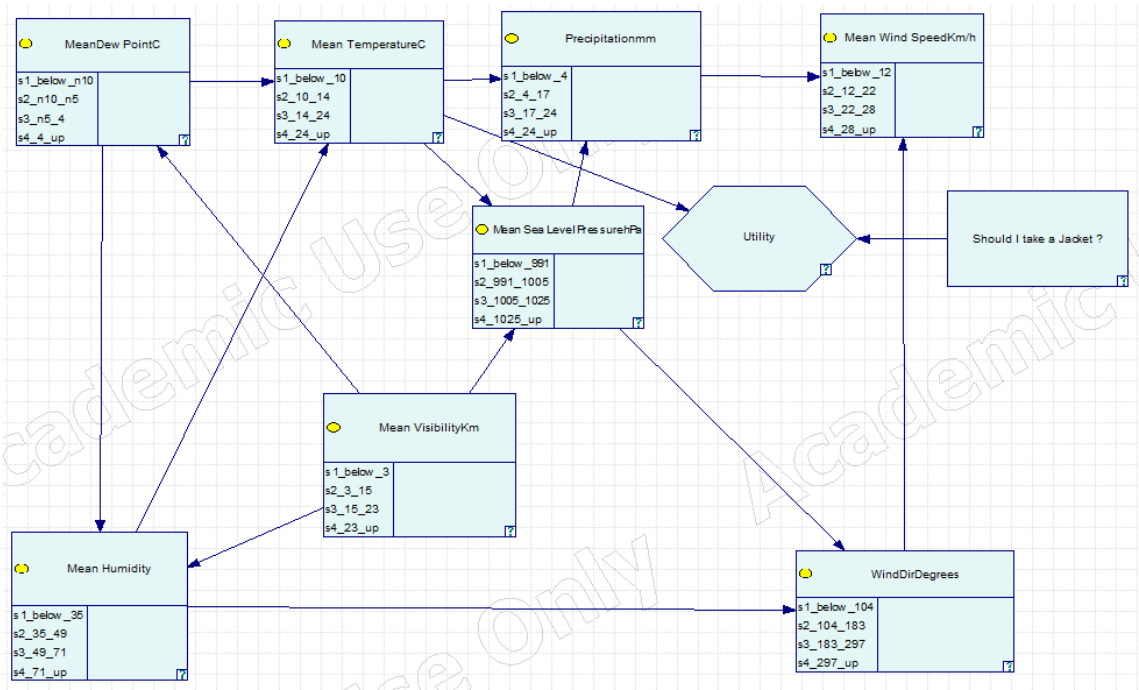
c. Query with Humidity < 35:



When the Mean Humidity is lower than 35, we notice changes in the temperature and in the Mean Visibility. This picture tells us that when we have low humidity, we have bigger temperatures and bigger visibility, which makes sense.

4.

In this part the goal is to add a decision node, my decision node is if I should take a jacket or not, the network with the addition is the following:



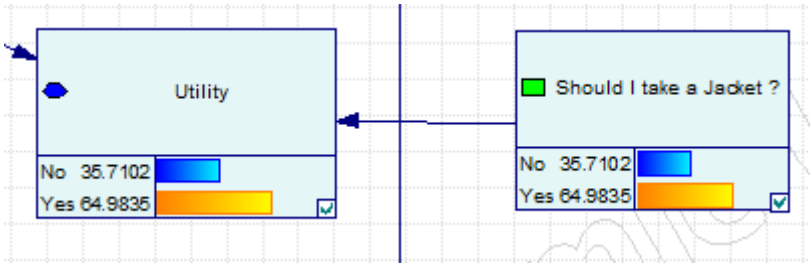
The utility of the node is the following:

Should I take a jacket?	No				Yes			
Mean Temperature	**< 10**	**10-14**	**14-24**	**>24**	**< 10**	**10-14**	**14-24**	**>24**
Value	0	20	50	100	100	70	30	5

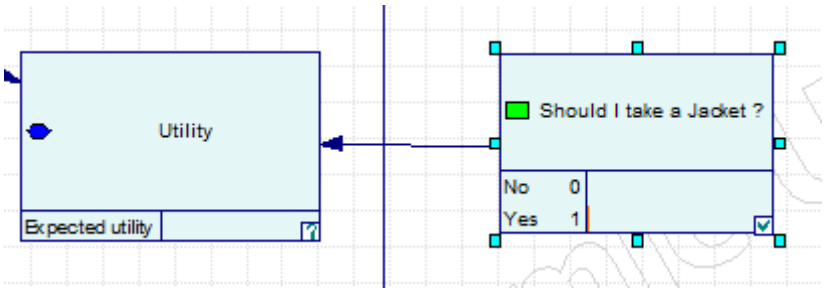
This table should have more variables, for example, Wind Speed I did only with temperature to keep it simple. To make this utility I thought about some facts:

- ➔ Temperature, if it's cold we must take the jacket.
- ➔ The fact that we can forget the jacket if it's too hot.

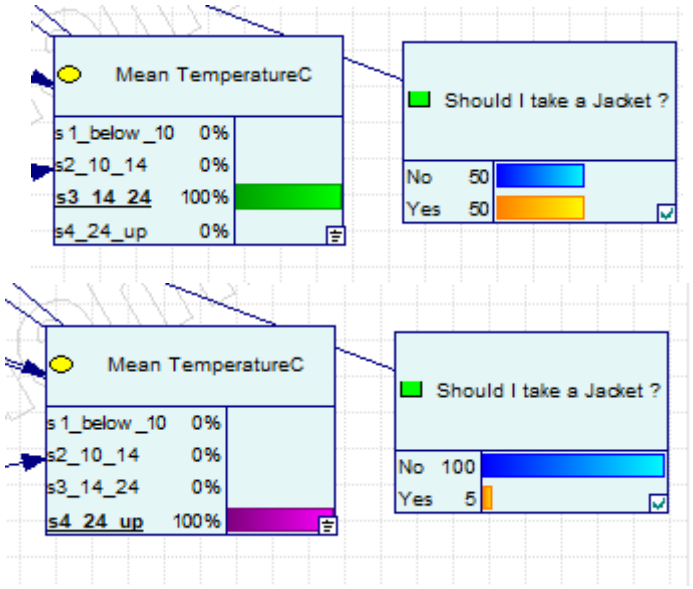
To solve this influences diagrams, I used the Policy Evaluation Algorithm to solve this. The Policy Evaluation Algorithm implementation is based on the algorithm proposed by Cooper (1998). The values obtained after running the algorithm and without any evidence are the following:



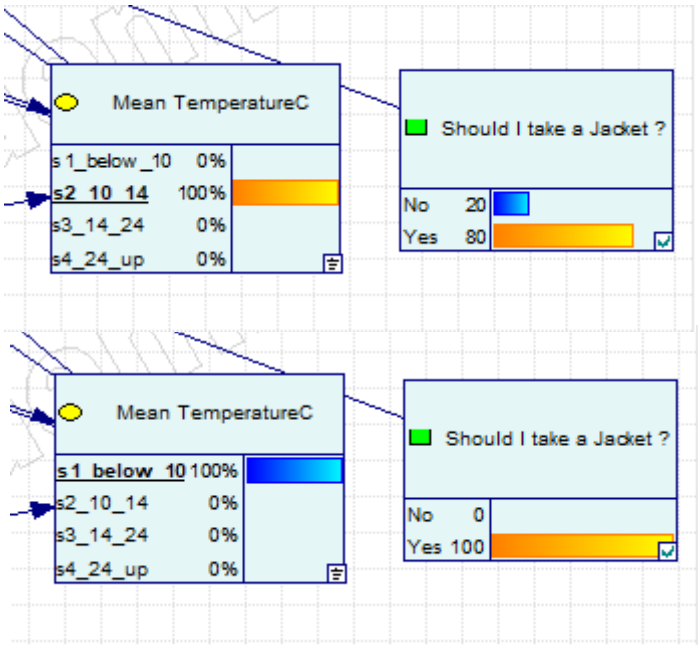
When I use the other algorithm called Find Best Policy, I get the following result:



If I put the Temperature > 14, I get the following result:

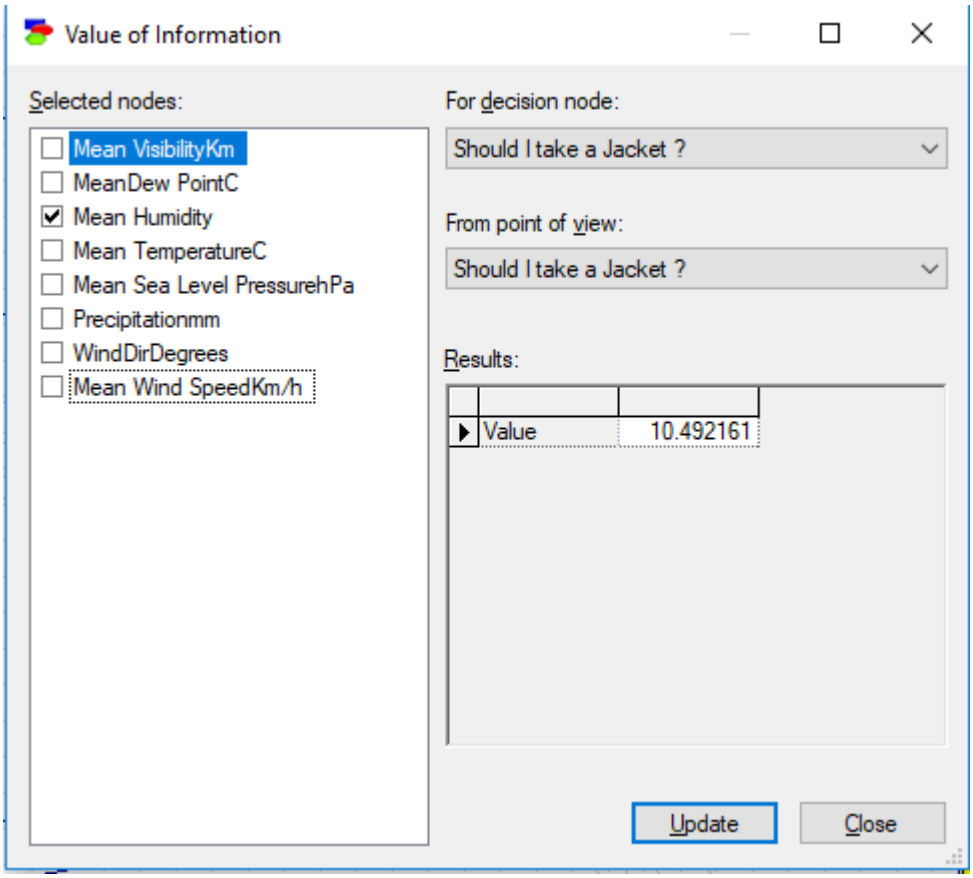


If I put the Temperature < 14, I get the following result:



5.

a) To compute the value of Value of Perfect Information (VPI) I used a tool in Genie called (Value of Information) and I choose the node Mean Humidity. I was able to get the following value:



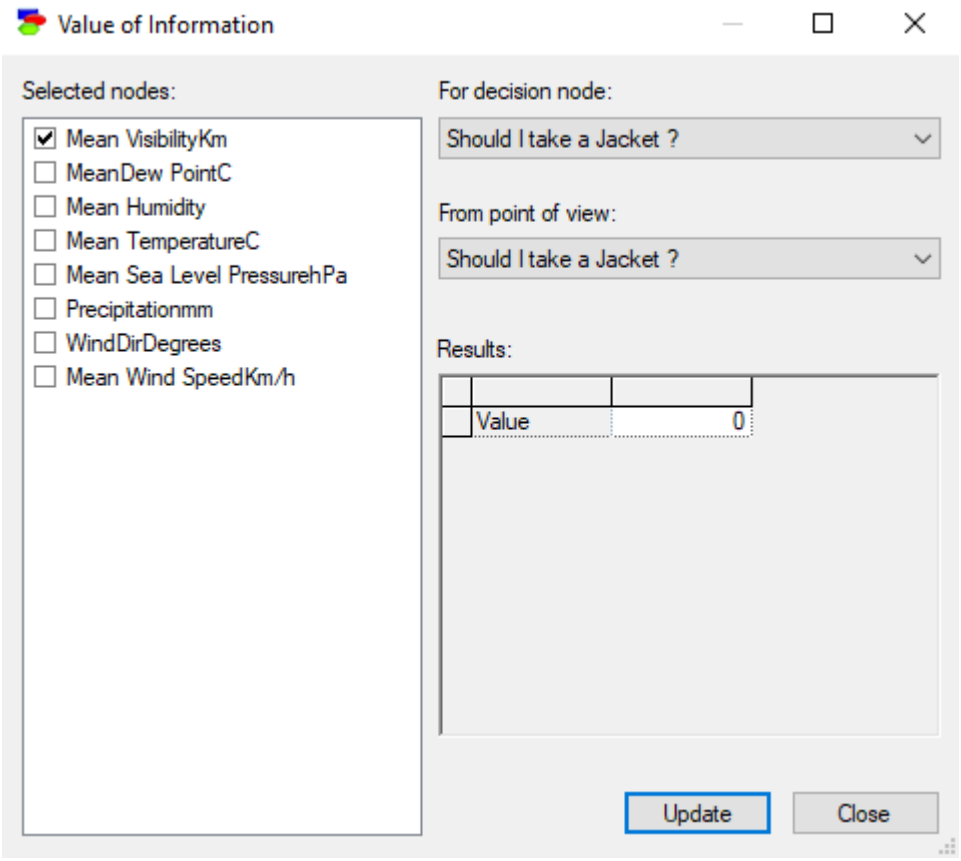
The value I obtained was 10.49. This value is the difference between the expected value of the optimal decision given perfect information about the Mean Humidity and the expected value of the optimal decision given no information about the node.

After connecting Mean Humidty to the decision node and after updating the values, we get the following result:

Should I take a jacket?	**No**				**Yes**			
Mean Humidity	**< 35**	**35-49**	**49-24**	**>71**	**< 35**	**35-49**	**49-24**	**>71**
Value	77.087	47.984	21.272	11.285	25.704	52.548	78.753	88.728

This says that the user should always take the jacket when the Mean Humidity is above 35. This means the user should only leave the jacket when we have a dry weather, otherwise, when the weather is wet the user should take the jacket.

b) When using the Value of Information tool, if we get any variable with the 0 value it means that this variable doesn't has an impact on the final value. This happens in the variable Mean Visibility.


Value of Information
Selected nodes:
Mean VisibilityKm
MeanDew PointC
Mean Humidity
Mean TemperatureC
Mean Sea Level PressurehPa
Precipitationmm
WindDirDegrees
Mean Wind SpeedKm/h
For decision node:
Should I take a Jacket ?
From point of view:
Should I take a Jacket ?
Results:
Value
0
Update
Close