A decision tree is the simple classifier for a person to classify if it will be cloudy or not tomorrow.

Feature selection:

The condition of being windy, cold, sunny and high pressure are the four factors affecting the decision of the class, all of which are based on the mentioned important variables. For the sake of simplicity of modelling, all predictor variables are categorical. Therefore, threshold values are set to determine if the level of those condition, which are listed below:

 Pressure is high if Pressure3pm and Pressure9am> 1013. (this is according to the international standard !!Cite)

It is cold when “MinTemp” < 10.

It is sunny when the hours of “sunshine” > 12.

It is windy if the “WindSpeed3pm” and “WindSpeed9am” > 15

The process of modelling:

1. 15 rows are sampled from the WAUS.Train for training the model
2. Calculate the entropy Initial State.



1.1 calculate the Information Gain for each predictor variables.



Table .1. Information gain: Cold



Table .2. Information gain: Pressure



Table .3. Information gain: Sunshine



Table .4 Information gain: Windy

From the tables above, Sunshine has the greatest information gain, thus becoming the root node.

* 1. Attribute to spilt on next

Gain(Not Sunny, Windy)=0.2212

Gain(Not Sunny, Cold)=0.0968

Gain(Not Sunny, Pressure)=0.0642

The above are the values of information gain for each variable given that the day is not sunny. Windy has the highest values, therefore it is the node when it is not sunny.

Gain(Sunny, Windy)=0.9710

Gain(Sunny, Cold)=0.9710

Gain(Sunny, Pressure)=0.9710

Likewise, the above are the values of information gain given that the day is sunny. Since one can observe that all the values are the same, therefore, the algorithm will select the feature randomly, in which Pressure is selected as a node when it is sunny.

* 1. 1.3 Attribute to spilt on next

Gain(Not Windy, Cold) = 0.4553

Gain( Windy, Cold) = 0.4200

Diagram

Description automatically generatedSince the value for Cold given that it is not windy is larger, the node of cold will be selected when it is not sunny and not windy.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Testing dataset** | | | | | |
| **Sunshine** | **Pressure** | **Windy** | **Cold** | **CloudTomorrow** | **Predicted Result** |
| Yes | Yes | Yes | No | Not Cloudy | Not Cloudy |
| No | Yes | Yes | Yes | Cloudy | Cloudy |
| No | Yes | Yes | Yes | Not Cloudy | Cloudy |
| No | No | No | No | Not Cloudy | Not Cloudy |
| No | No | Yes | No | Not Cloudy | Cloudy |

1. 5 rows are sampled from the WAUS.Test for test the model

The accuracy of the model = (1+2) / (5) = 0.6

Since this is made by hand