



**Module Code & Module Title**

**CU6051NA - Artificial Intelligence**

**Assignment**

**Year and Semester**

**2018-19 Spring**

**Student Name: Rajat Shrestha**

**London Met ID: 17030954**

**College ID: np01cp4a170021**

**Table of Contents**

[Decision Tree Assignment Solution 1](#_Toc27165270)

[Selecting the root node: 2](#_Toc27165271)

[Selecting nodes for Sunny and Rainy nodes: 5](#_Toc27165272)

[Building the final Decision tree: 7](#_Toc27165273)

**Table of Figures**

[Figure 1: Table of data 1](#_Toc27165315)

[Figure 2: Formulae to calculate Entropy and Gain 2](#_Toc27165316)

[Figure 3: Python function to calculate Entropy 2](#_Toc27165317)

[Figure 4: Python Function to calculate Gain 2](#_Toc27165318)

[Figure 5: Tally of each category of attributes 3](#_Toc27165319)

[Figure 6: Calculating the Gain value of each attributes 4](#_Toc27165320)

[Figure 7: Tally of each category on Sunny node 5](#_Toc27165321)

[Figure 8:Tally of each category on Rainy node 5](#_Toc27165322)

[Figure 9: Gain calculated in Sunny node 6](#_Toc27165323)

[Figure 10: Gain Calculated in Rainy node 6](#_Toc27165324)

[Figure 11: Final decision tree 7](file:///C:\Users\Rajat\Desktop\Rajat%20Shrestha%20Desision%20Tree.docx#_Toc27165325)

# Decision Tree Assignment Solution

Build a decision tree using data from the given table. The data is about whether to play tennis or not given the weather conditions. So, when the tree is built, it should be able to give out a decision on whether to play tennis or not given the weather conditions.

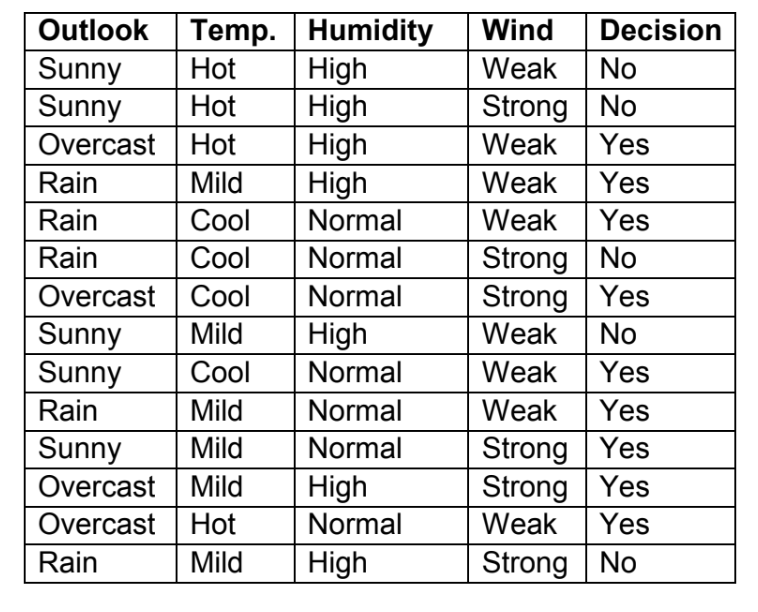


Figure 1: Table of data

To build a decision tree we must go through various process to find the nodes:

## Selecting the root node:

Here, we have four different features which is given the final decision to construct a decision tree we must first find the root node. To choose the optimal root node we must find the attribute with the maximum amount of gain.

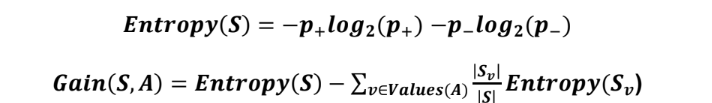


Figure 2: Formulae to calculate Entropy and Gain

Since we need to do this calculation multiple times a function was created to take the count of positive and negative outcomes and give the entropy as well as another function to take the overall positive and negative outcomes with the list of all the positive and negative outcomes of individual category of an attribute:

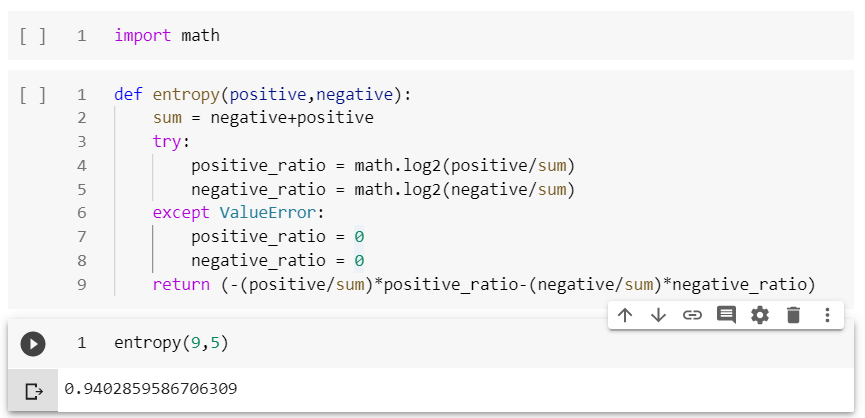


Figure 3: Python function to calculate Entropy

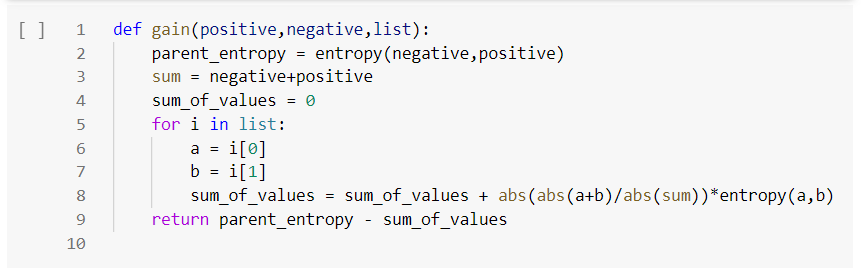


Figure 4: Python Function to calculate Gain

Now, to calculate the gain of each column we need to calculate the entropy of the all the labels categorized in an attribute, we must tally down all the positive and negative outcomes of all existing categories of the attributes.

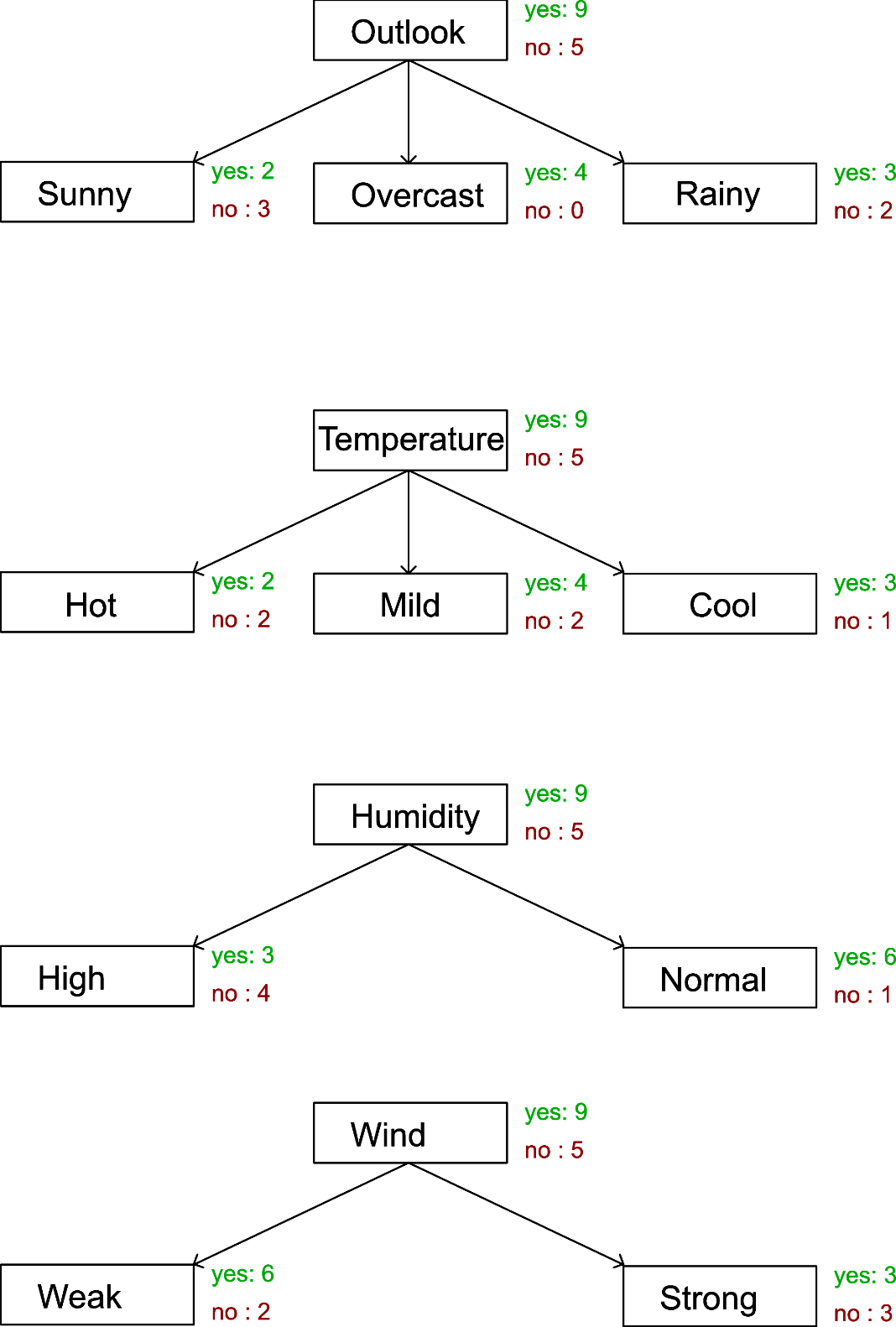


Figure 5: Tally of each category of attributes

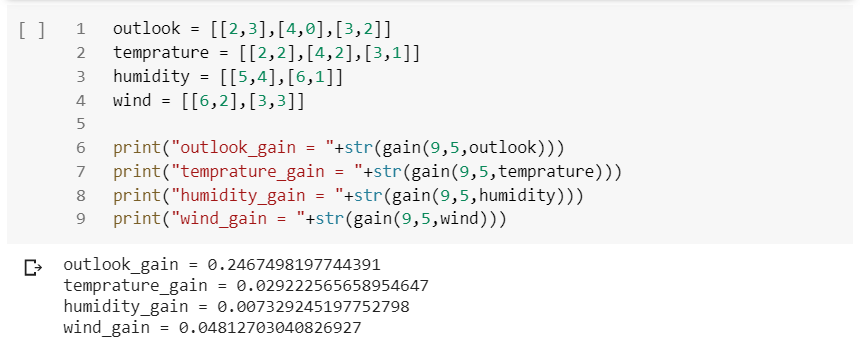


Figure 6: Storing above data in lists and calculating the Gain value of each attributes

As we can see the highest gain value is of outlook so taking outlook as the root node.

Now, as we got the root node as Outlook, observing the three root nodes, overcast is totally saturated while sunny and rainy is not. So, at the next step the same procedure is carried out with stripped data including the selected category.

## **Selecting nodes for Sunny and Rainy nodes:**

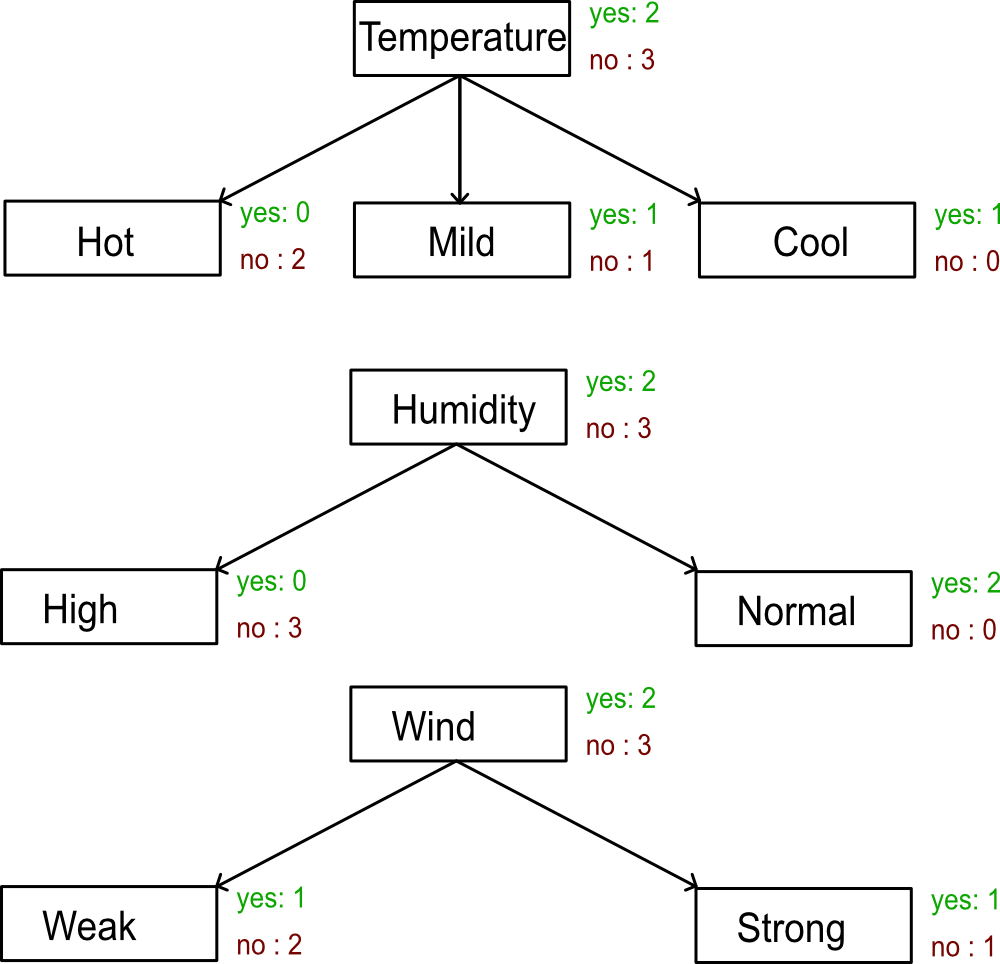


Figure 7: Tally of each category on Sunny node

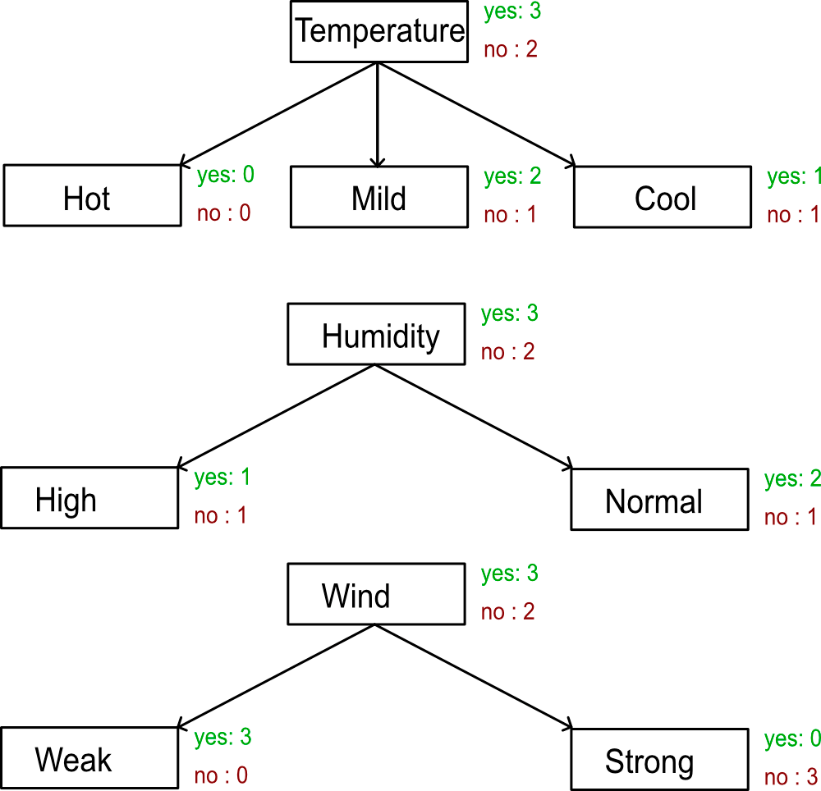


Figure 8:Tally of each category on Rainy node

Calculations using above data:

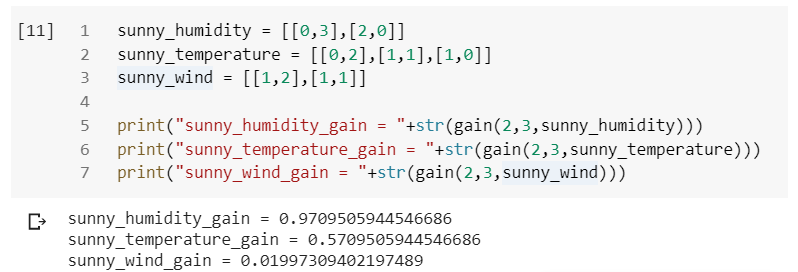


Figure 9: Gain calculated in Sunny node

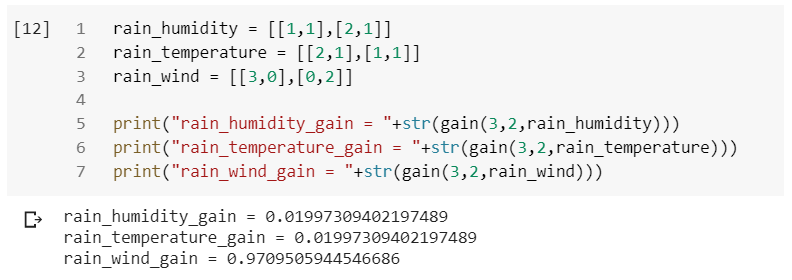


Figure 10: Gain Calculated in Rainy node

As we can see in the sunny node, humidity is the node with lowest saturation 0 and the highest gain and in the rainy node, wind gives the lowest degree of saturation and highest gain value so taking them as the child nodes.

Note: these calculations were unnecessary but if the nodes were not saturated it should be done to get the node with highest gain value.

## Building the final Decision tree:

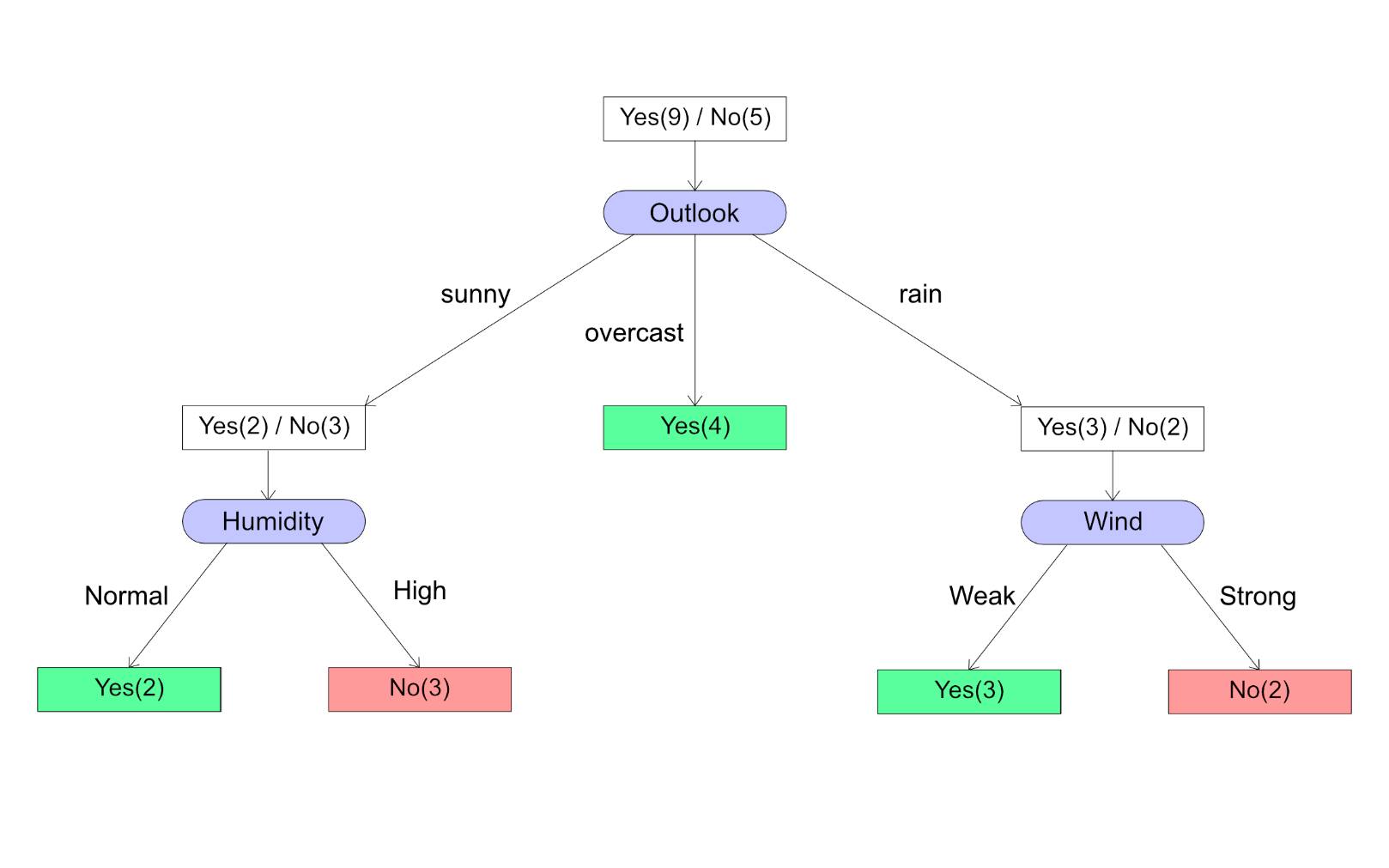
Finally, from the above calculations and observations, the resulting decision tree is as follows:

Figure 11: Final decision tree

The link to the notebook used to calculate the entropy and gain(a):

<https://colab.research.google.com/drive/1fj8nGbi0hC6d0_8mUww1Ea14LHhi_X4Q>