Machine Learning Guess Happiness

Recommendation: Let’s use the Decision Tree Model from Class #2 activities 3-4

## Data Preparation Requirement: just a good representation of the problem.

**Data**

World Happiness Report Country and Rank (2019)

Carl Jungs 5 Factors of Happiness

**Interaction**

Country – get rank

Calculate Score :

* Good Physical Health
* Good persona and intimate relationship (marriage family friends)
* The faculty to perceive beauty in art and nature
* Reasonable stands of living/satisfactory work
* Philosophic or religious point of view capable for coping successfully with the vicissitudes of life

**Score Calculation:**

Responses: Yes No It’s Complicated

Value: Yes = 10 No = 1 Complicated = 5

Output – happiness emoticon

**Logic** Location \* Value Weight = Happiness

Happiness

### **Output Decision Tree**

|  |  |  |
| --- | --- | --- |
| Result – use emoticon happy faces | Range (Country) | CJ Factor Total\* |
| Very Happy | Ranks 1-10 | 45-50 |
| Happy | Rank 11-50 | 35-44 |
| Somewhat Happy | Rank 75-125 | 30-34 |
| Not Happy | Rank >125 | < 30 |

\*Any combination of the CJ responses that total within the range

Additional ML Research Support Approach:

Data Source for the following: <https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/>

Machine Learning Algorithm Type:

Recommended: Classification Supervised Learning Algorithm

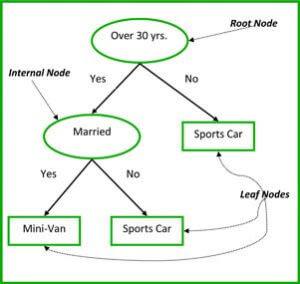
**Classification** is used to predict the outcome of a given sample when the output variable is in the form of categories. A classification model might look at the input data and try to predict labels like “sick” or “healthy.”

Recommended Model: CART

**Classification and Regression Trees (CART)** are one implementation of Decision Trees.

The non-terminal nodes of Classification and Regression Trees are the root node and the internal node. The terminal nodes are the leaf nodes. Each non-terminal node represents a single input variable (x) and a splitting point on that variable; the leaf nodes represent the output variable (y). The model is used as follows to make predictions: walk the splits of the tree to arrive at a leaf node and output the value present at the leaf node.

The decision tree in Figure 3 below classifies whether a person will buy a sports car or a minivan depending on their age and marital status. If the person is over 30 years and is not married, we walk the tree as follows : ‘over 30 years?’ -> yes -> ’married?’ -> no. Hence, the model outputs a sports car.



## Data Preparation for CART

CART does not require any special data preparation other than a good representation of the problem. <https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/>

STEP-BY-STEP Example

<https://sefiks.com/2018/08/27/a-step-by-step-cart-decision-tree-example/>

### Output Decision Tree

Happiness

|  |  |  |  |
| --- | --- | --- | --- |
|  | Country Rank | CJ Score\* | Decision |
| 1 | 1-2 | 45-50 | Very Happy |
| 2 | 3-4 | 35-44 | Happy |
| 3 | 5-6 | 30-34 | Somewhat Happy |
| 4 | 7-8 | < 30 | Not Happy |

\*CJ Score – various combination of responses that equate to total

Other ML Models reviewed:

1. <https://shiring.github.io/machine_learning/2017/04/23/one_r>
2. <https://rstudio-pubs-static.s3.amazonaws.com/201826_cab699be72ca47f99debadf16ee54c95.html>
3. <https://dl7631.shinyapps.io/countries_happiness/>
4. https://nycdatascience.com/blog/r/increasing-countries-happiness/