Modeling with Decision Trees

# Decision Tree Definition

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm.

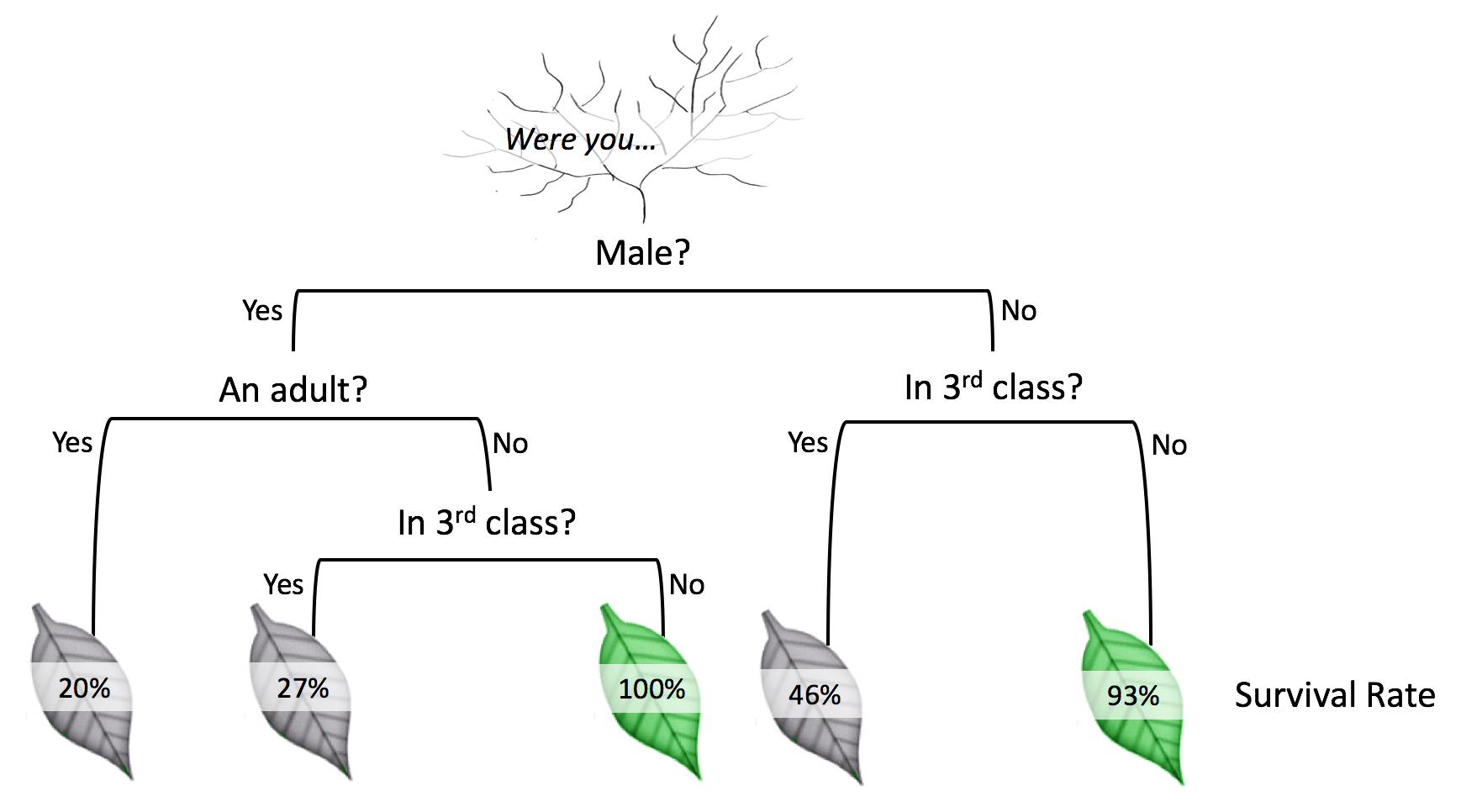


Figure 1 - Decision tree example

# Decision Tree Usage

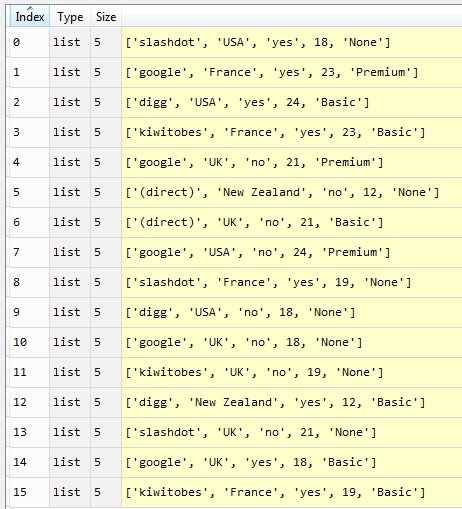
Decision trees are used to solve regression and classification problems. *Classification trees*, are used to separate the dataset into classes belonging to the response variable. Usually the response variable has two classes: Yes or No (1 or 0). Thus classification trees are used when the response or target variable is categorical in nature.

*Regression trees* are needed when the response variable is numeric or continuous. For example, the predicted price of a consumer good. Thus, regression trees are applicable for prediction type of problems as opposed to classification.

In the Final Project, the task is to predict whether users will sign up for a subscription based service. This is *classification* problem since only “None”, “Basic” and “Premium” outcomes are possible.

# Decision algorithm

* Step 1: Define a list of features. In our case they are:
  + The site that referred them
  + Geographical location
  + How many page views
  + Service chosen



* Step 2: Define dividing function to split a set into 2 subsets, according to criteria. The algorithm uses this function for four criteria: site, location, FAQ, number of pages.
* Step 3: Define entropy function (an objective criterion to split a group). This function is used after division with different criterion. Based on the function value the best possible division criterion is selected.
* Step 4: In loop for every criterion (column):
  + execute division into two sets
  + calculate objective criterion (entropy function)
  + choose the best result
* Step 5: After first split is done recursion is used for further splits

Finally, decision tree is built. After that the tree is drawn:

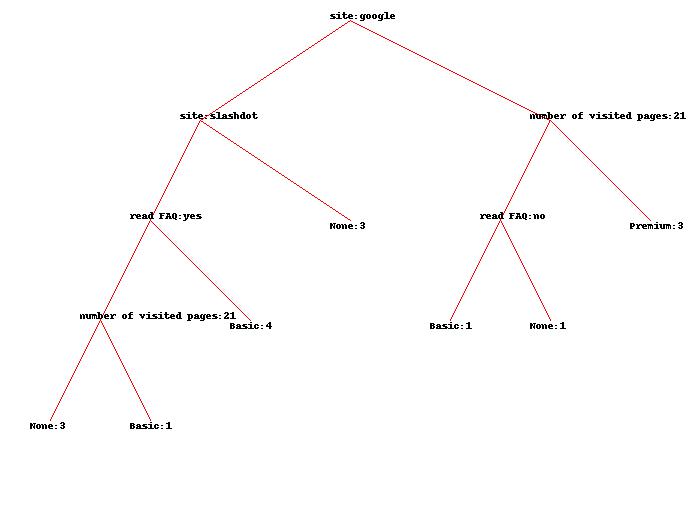


Figure 2 - Final decision tree