NexTint Data Demo Decision Tree Using RPart

In June 2015, the startup, NexTint, was struggling with trying to find the factor that produced the most stable devices. The following is an analysis that I did on all of the devices ever made at NexTint to find the most important factor governing stability. The obvious factors were anode material, cathode material, and electrolyte material; and, I also checked the relationship between the fabrication date and long term stability.

I was curious to find the most important factor for determining whether or not the devices on our long term stability tester lasted more than one month. Figure 1 shows the decision tree classifier that was produced using the script NexTintTree.R. NexTintTree.R uses the cleandf dataframe produced in frame 8 of the IPython notebook located at

https://github.com/tobysachsquintana/data_science_portfolio/blob/master/NexTint%20Data%20Demo.ipynb

What is very interesting and very surprising is that the tree for devices with a lifetime greater than a month, never splits on any other parameter besides date. The first split is on month >= 1.5 (month >= Nov. 2014).

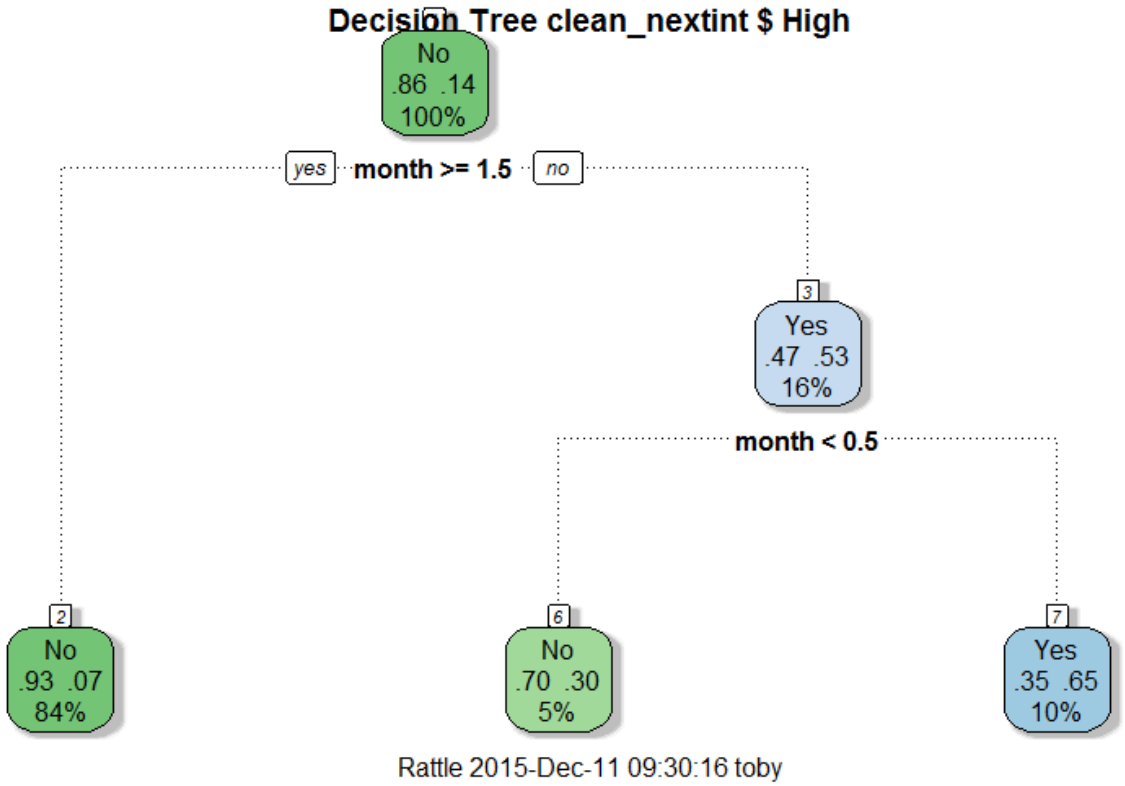

Figure 1: Decision tree classifier based on whether or not a device spent more than a month on the lifetime tester. The script for the decision tree can be found at https://github.com/tobysachsquintana/data_science_portfolio/blob/master/NexTintTree.R