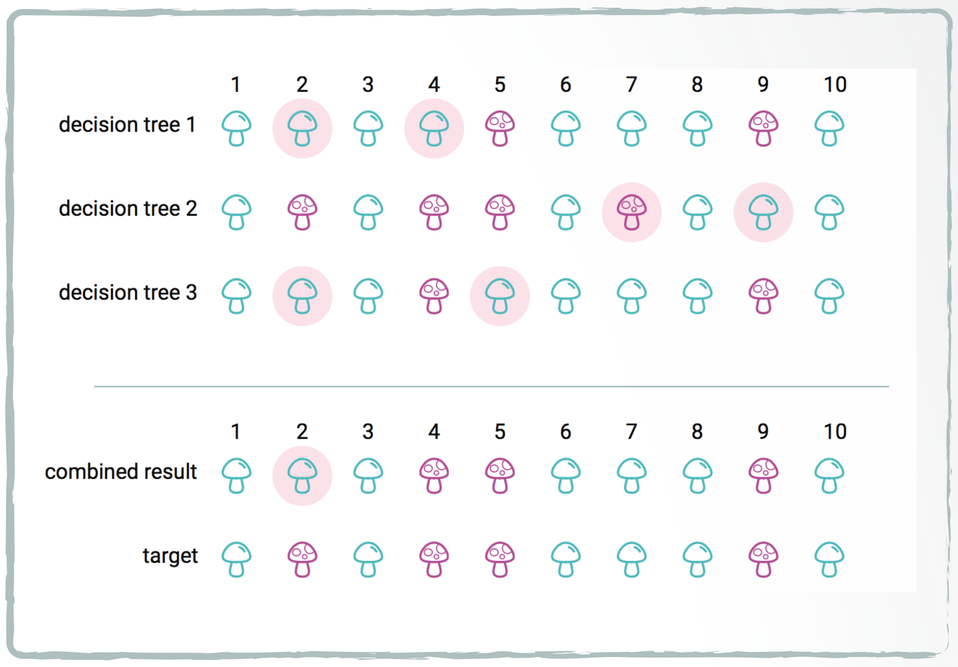
A Path to the Random Forest

Machine learning algorithm visualization using a mushroom edibility classification example

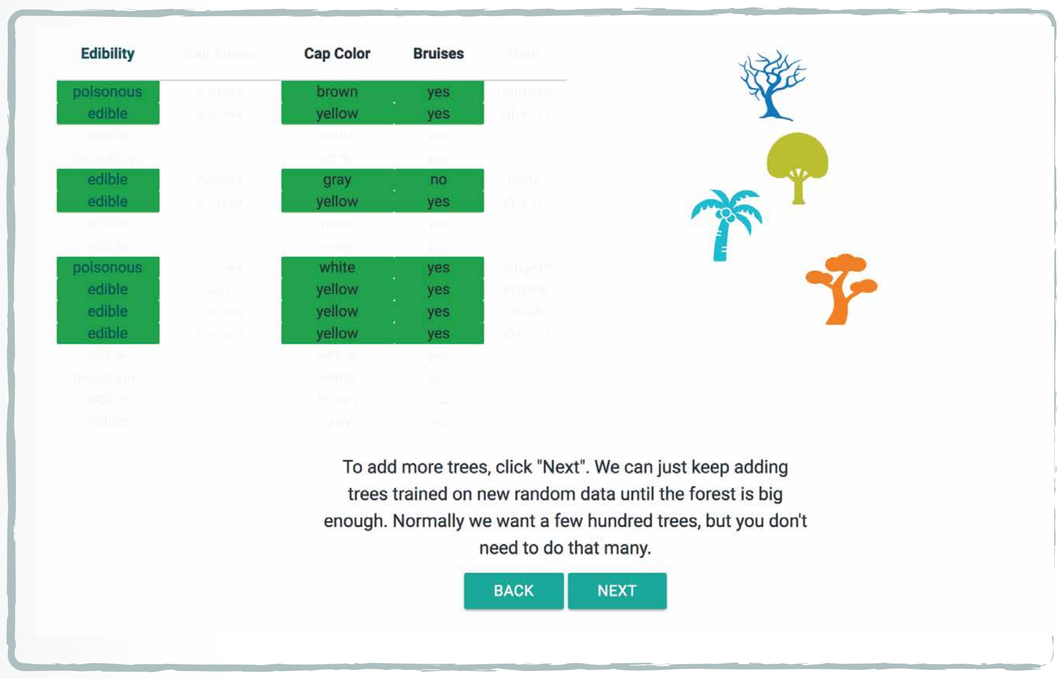
Fallacy of Individual Predictors

This static visualization explains ensemble predictors. Easily iconified mushroom shapes are memorable while allowing the reader to focus on the accuracy of the ensemble predictor.



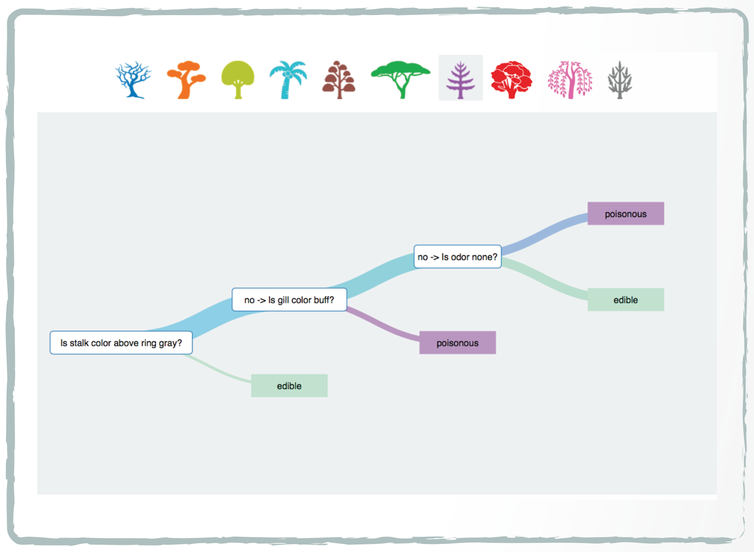
Training of Trees

Readers step through an animation that explains how data and features are selected for each tree in the forest. Crucially, users can go back and review previous steps. This control can improve learning by increasing the readers' interactions. Also, each step has a text explanation to resolve any confusion.



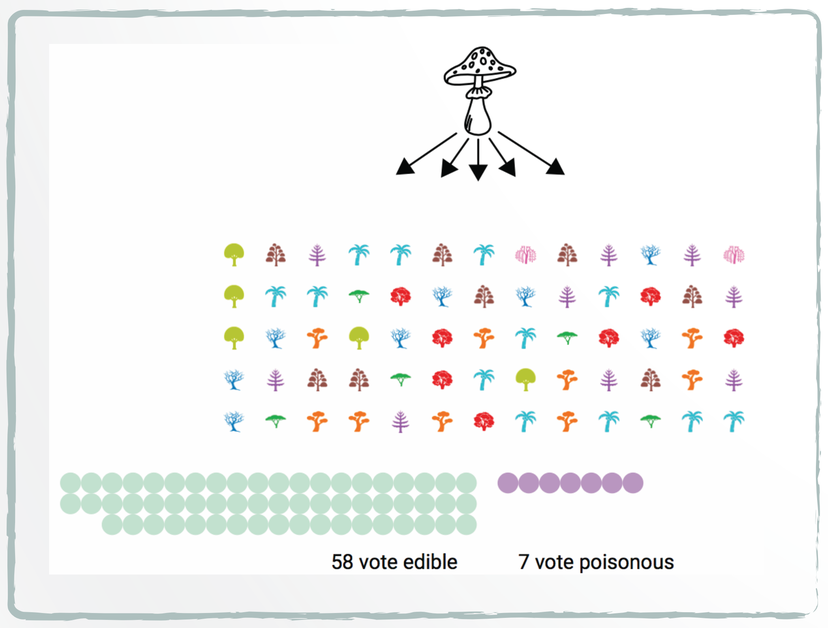
Training Result

Each decision tree trained in the previous step can be inspected in this view. The same decision tree icons used in the previous step are visible here. Readers interactivily click on the decision nodes to expand and collapse branches. The green and purple coloring of safe and poisonous mushrooms is consistent with icons used earlier in the design.



Making a New Prediction

Finally, the visualization shows how decision trees in a random forest work together to decide whether a new mushroom is poisonous or not. The animation shows the process of majority voting in a dynamic, engaging and easy to follow fashion. As before, users move forward and back and icons and colors are consistent with previous illustrations.



Principles for Machine Learning Algorithm Visualization

1. Create an approachable narrative. Choose an example data set that aids the story and use memorable icons.
2. Ensure all steps are controllable and reversible by the user. Explain the meaning of each animation.
3. Use visual components (color, iconography, gestalt properties) consistently throughout the design.