

Al Shafian Bari

December 16, 2023

Decision Tree

The build of a decision tree starts with a root node, which does not have any incoming branches. That is the top node in the tree where it starts the test. Then we have the decision nodes which are also called the internal nodes. This is where the test of the specific features is taking place and the decision at each node leads to a different path based on the test. Next, we have the branches that make the tree. The branches are used to connect each node together and the different outcomes to come. Lastly, we have the leaf nodes. This is the final node on the tree and the overall final decision of the whole tree.

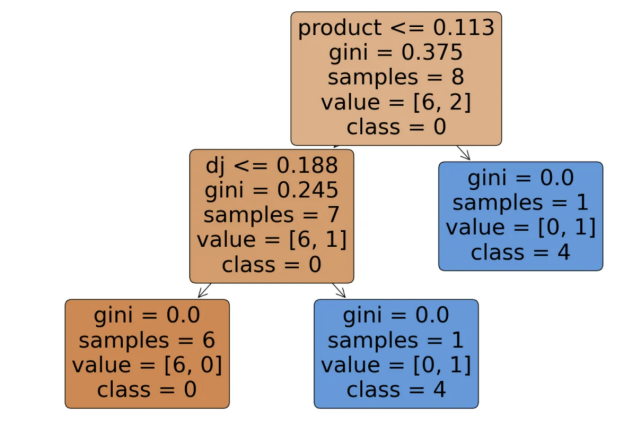
There are steps that decision trees have to go through in order to get to the final decision. One of these steps is selecting the best features. The algorithm checks the different features and then selects the one that provides the best split. When splitting the decision the tree uses the dataset that was given and observes the overall features. For example, if there is a dataset that has to do with age the decision tree might split them up into ages greater than forty-five and ages less than forty-five.

The outputs of the code provide a comprehensive overview of the decision tree model's performance on a randomly selected subset of 10 rows from the dataset. The accuracy on the subset reveals the percentage of correctly predicted instances, giving you an initial impression of how well the model generalizes to this specific portion of the data. The confusion matrix takes the evaluation a step further by breaking down predictions into true positives, true negatives, false positives, and false negatives. This matrix offers insights into specific areas where the model excels or struggles, aiding in the identification of potential biases or areas for

improvement. Additionally, the display of 10 random rows alongside their actual and predicted values allows for a qualitative assessment of the model's predictions on individual instances, offering a closer look at its performance at a granular level.

Accuracy on subset: 100.00%
Confusion Matrix on subset:
[[2]]

	Text	Target	Predicted
541200	@chrishasboobs AHHA I HOPE YOUR OK!!!	0	0
750	@misstoriblack cool , i have no tweet apps fo...	0	0
766711	@TiannaChaos i know just family drama. its la...	0	0
285055	School email won't open and I have geography ...	0	0
705995	upper airways problem	0	0
379611	Going to miss Pastor's sermon on Faith...	0	0
1189018	on lunch....dj should come eat with me	4	4
667030	@piginthepoke oh why are you feeling like that?	0	0
93541	gahh noo!peyton needs to live!this is horrible	0	0
1097326	@mrstessyman thank you glad you like it! There...	4	4



The decision tree plot visually represents the decision-making process of the model, illustrating the hierarchy of features and their importance in predicting the target variable. Each node in the tree represents a decision based on a specific feature, and the branches depict possible outcomes. The color and size of nodes convey information about the predicted class and the number of samples, respectively.