Decision Tree Variables and Design

In the golf example, each outcome is independent in that it does not depend on what happened in the previous coin toss. Dependent variables, on the other hand, are those that are influenced by events before them.

Building a decision tree involves construction, in which you select the attributes and conditions that will produce the tree. Then, the tree is pruned to remove irrelevant branches that could inhibit accuracy. Pruning involves spotting outliers, data points far outside the norm, that could throw off the calculations by giving too much weight to rare occurrences in the data.

Maybe temperature is not important when it comes to your golf score or there was a day when you scored poorly that’s throwing off your decision tree. As you’re exploring the data for your decision tree, you can prune specific outliers like your one bad day on the course. You can also prune entire decision nodes, like temperature, that may be irrelevant to classifying your data.

Well-designed decision trees present data with few nodes and branches. You can draw a simple decision tree by hand on a piece of paper or a whiteboard. More complex problems, however, require the use of decision tree software.

Types of Decision Trees

There are two main types of decision treesExternal link: open\_in\_new: categorical and continuous. The divisions are based on the type of outcome variables used.

Categorical Variable Decision Tree

In a categorical variable decision tree, the answer neatly fits into one category or another. Was the coin toss heads or tails? Is the animal a reptile or mammal? In this type of decision tree, data is placed into a single category based on the decisions at the nodes throughout the tree.

Continuous Variable Decision Tree

A continuous variable decision tree is one where there is not a simple yes or no answer. It’s also known as a regression tree because the decision or outcome variable depends on other decisions farther up the tree or the type of choice involved in the decision.

The benefit of a continuous variable decision tree is that the outcome can be predicted based on multiple variables rather than on a single variable as in a categorical variable decision tree. Continuous variable decision trees are used to create predictions. The system can be used for both linear and non-linear relationships if the correct algorithm is selected.

What Are the Applications for a Decision Tree?

Decision trees are useful for categorizing results where attributes can be sorted against known criteria to determine the final category.

Decision trees map possible outcomes of a series of related choices. Some applications for decision trees include:

1.Customer Recommendation Engines

Customers buying certain products or categories of products might be inclined to buy something similar to what they’re looking for. That’s where recommendation engines come in. They can be used to push the sale of snow gloves with a purchase of skis or recommending another holiday movie after you’ve just finished watching one.

Recommendation engines can be structured using decision trees, taking the decisions made by consumers over time and creating nodes based off of those decisions.

2.Identifying Risk Factors for Depression

A study conducted in 2009 in Australia tracked a cohort of over 6,000 people and whether or not they had a major depressive disorder over a four-year period. The researchers took inputs like tobacco use, alcohol use, employment status and more to create a decision tree that could be used to predict the risk of a major depressive disorder.

Medical decisions and diagnoses rely on multiple inputs to understand a patient and what may be the best way to treat them. A decision tree application like this can be a valuable tool to health care providers when assessing patients.