

Decision Tree

1. Suppose there is an attribute, "A," that consists of random values, and these values do not have any correlation with the class labels. Additionally, assume that "A" has a sufficient number of distinct values such that no two instances in the training dataset share the same value for "A." What would be the outcome if a decision tree is built using this attribute? What challenges or issues might arise in this scenario?

2. Answer the questions according to the following dataset:

Weekend	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

- a) Create a decision tree model using the given dataset to predict the value of the final column, using all other columns as input features except for the first one(weekend). Clearly explain each step of the process, including your calculations, reasoning, and decisions made while constructing the tree. What is the model's overall classification accuracy?
- b) Can you design a specific set of training examples that would cause the algorithm to include "Weather" as an attribute in the final decision tree, even if the actual target concept is not dependent on "Weather"?if not,explain why .if yes,provides such a set of examples.
- c) Construct a decision tree model using only the first 6 samples from the dataset(W1_W6). Evaluate the model's classification performance on these initial 6 samples as the training set. Then, use the model to classify the remaining samples in the dataset. What is the classification accuracy for both the training and test datasets? Discuss your findings and explain the reasons behind the observed results.
- d) In scenarios where only a limited number of labeled examples are available for training (and no extra data is available for testing or validation), propose a specific pruning technique that could be integrated into the decision tree algorithm to prevent overfitting. Justify why you believe this technique would be effective.