

Part 6

EXPLAIN HOW DECISION TREE IS SUITABLE HANDLE MISSING VALUES IN DATA.

There are several ways to handle missing values –

- Fill the missing value with mean or median of the values if data is numerical.
- If data is categorical you can fill missing values with the value that has appeared maximum in that label.
- Fill data with the values that come maximum in the range below and above that value.
- For the node in the tree whose value is missing in test data , you check both braches and count the number of times each possible outcome label appears at the leaf and then return the label which appears most.
- Deleting the row simplest method is to delete the entire row where data is missing. If data has high variance then this lose of data is harmful.