

DECISION TREES

- Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.
- They are simple to understand and to interpret. Trees can be visualised. Requires little data preparation. Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed.
- Able to handle both numerical and categorical data. Able to handle multi-output problems.
- Uses a white box model. If a given situation is observable in a model, the explanation for the condition is easily explained by boolean logic. By contrast, in a black box model (e.g., in an artificial neural network), results may be more difficult to interpret.
- Possible to validate a model using statistical tests. That makes it possible to account for the reliability of the model.
- Performs well even if its assumptions are somewhat violated by the true model from which the data were generated.
- However, decision trees can be unstable because of small variations in the data which might result in a completely different tree being generated. This problem is mitigated by using decision trees within an ensemble.
- `DecisionTreeClassifier` is a class capable of performing multi-class classification on a dataset.

- An example of decision tree visualization:

