

## LAB-2

### ID3 Algorithm.

1. Start with entire dataset.

\* the goal is to construct a decision tree by recursively splitting the dataset based on the features that provide the most information.

2. check if the dataset is pure

\* If all instances in the dataset have the same class label, return that class as the result (this is a leaf node).

3. check if the dataset is empty

\* If the dataset is empty, return the most frequent class from the parent node.

4. check if there are no more features to split

\* If there are no remaining features, return the most frequent class in the current dataset (this is a leaf node).

5. Calculate Information Gain

\* for each feature in the dataset, calculate the information gain based on entropy

$$IG(D, A) = \text{Entropy}(D) - \sum_{\text{value}(A)} \left( \frac{|D_v|}{|D|} \right) \text{Entropy}(D_v)$$

$D$  - dataset,  $A$  - feature,  $D_v$  - subset of  $D$

$\text{value}(A)$  :- unique values of feature  $A$ .

6. choose the feature with highest IG.

\* the feature with the highest information gain is selected as feature and splitting criterion.



7. split the dataset based on the selected feature
  - \* Create a branch for each unique value of the selected feature and partition the dataset into subsets for each branch.
8. Recursively apply ID3 to each subset
  - \* Recursively apply the ID3 algorithm to each subset of the dataset until one of the stopping conditions is met.

9. Prune the tree

In some cases, after the tree is built, it can be pruned by removing branches that don't add value, typically using data.

date 10/3/24