— : LAB : 02 :— Date 10/03/25 Page 1

Exploring ID3 Algorithm :-

Mainly Algorithm is used for,
- classification
- Identification
- Prediction
- Data Information etc.
- Building decision tree etc.

A Recurssive algorithm, used under the supervised learning.

Notes to be Understand :

Decision Tree :-

a flowchart-like structure where each internal node reprents a feature (attribute) in the dataset, each branch represents a decision rule, and each leaf node represents a class label.

Entropy :-

a measure of uncertainity or impurity of a dataset.

It quantifies how ~~tired~~ mixed the dataset is in terms of its class labels.

$$\text{Entropy}(S) = -\sum_{i=1}^{k} P(c_i) \log_2 P(c_i)$$

- S → dataset
- c_i → class label
- $P(c_i)$ → probability of class c_i in dataset S

Date / /
Page

* Information Gain (IG) :-
measures how well a feature separates the dataset into different classes. The higher the IG, the better the feature is at classifying the examples.

So, basically IG created by comparing the entropy of the dataset before and after splitting the dataset based on a particular feature.

$$IG(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

A -> attribute (feature),

=> Steps in ID3 Algorithm :-

1. Choose the best attribute
2. Split the dataset
3. Create a new node
4. Repeat the process recursively
5. Apply stopping condition.

For Example :

* Calculation of Entropy of entire dataset.
9 times Yes and
5 times No

$$(S) = -\left(\frac{9}{14}\log_2\frac{9}{14} + \frac{5}{14}\log_2\frac{5}{14}\right)$$

$$S \approx 0.940$$

Decision Tree :-