**Decision Trees (ID3, CART)**

**Decision Tree Introduction:** A Decision Tree is a supervised machine learning algorithm used for both classification and regression tasks. It models decisions in a tree-like structure of nodes and branches:

* **Root Node:** The starting point of the tree (entire dataset).
* **Decision Nodes:** Nodes where data is split based on a feature.
* **Leaf Nodes:** Terminal nodes that provide the final prediction (class label or value).

**Working:** It splits the dataset into subsets based on the feature that results in the **best split** using metrics like **Entropy, Gini, or MSE**.

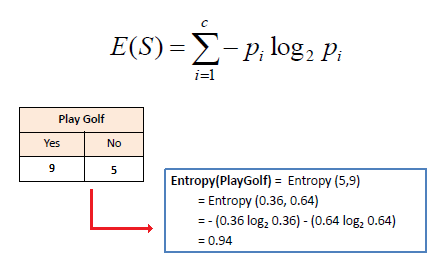
**Types of Decision Trees:**

* ID3
* CART

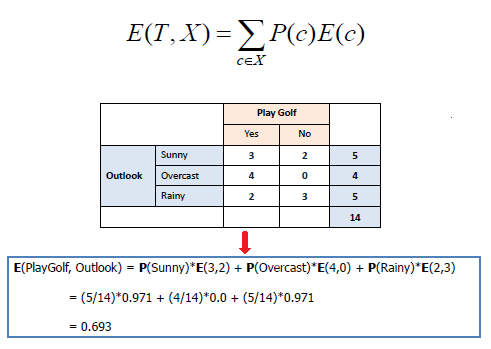
**ID3:** Works as follows:

* Calculate entropy for the dataset.
* For each feature, calculate information gain.
* Select the feature with the highest information gain to split.
* Repeat recursively on subsets until stopping criteria (all same class or no more features).

**Entropy Formula:**

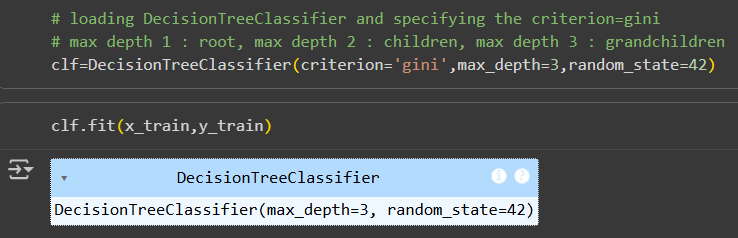


**Information Gain Formula:**

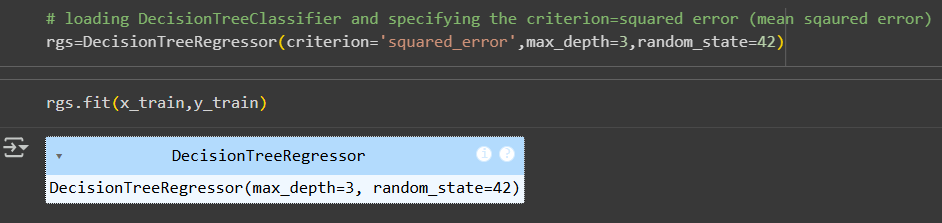


**CART:** Two types of CART:

* **Classification (CART):**
* Input: Features (numeric or categorical).
* Output: Class label (e.g., “spam” or “ham”, “disease” or “no disease”).
* Split criterion: Gini Index or Entropy.



* **Regression (CART):**
* Input: Features (numeric or categorical).
* Output: Continuous value (e.g., house price = 250,000).
* Split criterion: MSE (Mean Squared Error) or MAE (Mean Absolute Error).



**Ensemble Intro & Bagging**

**Ensemble Learning:** Combining multiple models (weak learners) to build a stronger, more accurate model. There are multiple ensemble learning methods/models:

* **Bagging**
* **Boosting**
* **Stacking**

**Bagging (Bootstrap Aggregating):** Bagging is one of the most common ensemble techniques. It works in the following way:

* **Bootstrapping**

From your dataset, create many **random samples with replacement**.

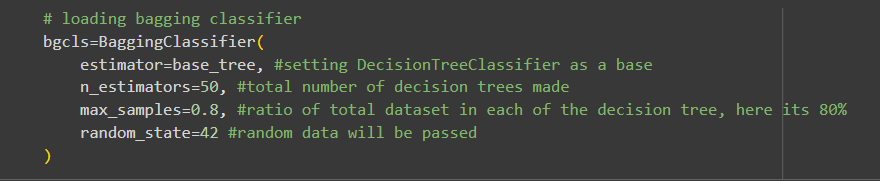
* Each sample is slightly different.
* **Train models**

Train one model (e.g., a decision tree) on each bootstrap sample.

* **Aggregate results**

For classification → take a **majority vote**.

For regression → take the **average prediction**.



**Random Forest**

**Random Forest:** Ensemble of Decision Trees

* It uses **bagging** (bootstrap aggregating)
* Plus, **random feature selection** at each split.

So instead of growing **one tree** (which might overfit), we grow **many trees** and combine them.

**Working:** Ensemble of Decision Trees

* Draw random samples of data (with replacement).
* Train a Decision Tree on each sample.

At each split, it only considers a random subset of features (not all).

* Aggregate results:

For classification → **majority vote**.

For regression → **average prediction**.

**Comparison with a Single Decision Tree:**

* A single tree has **high variance** (changes a lot with small changes in data).
* Random Forest reduces variance by averaging many trees.
* More stable, more accurate, less overfitting.

**Key Hyperparameters:**

* **n\_estimators:** number of trees (default 100, more = better but slower).
* **max\_features:** how many features to consider at each split.
* **max\_depth:** maximum depth of each tree.
* **min\_samples\_split, min\_samples\_leaf:** control overfitting.