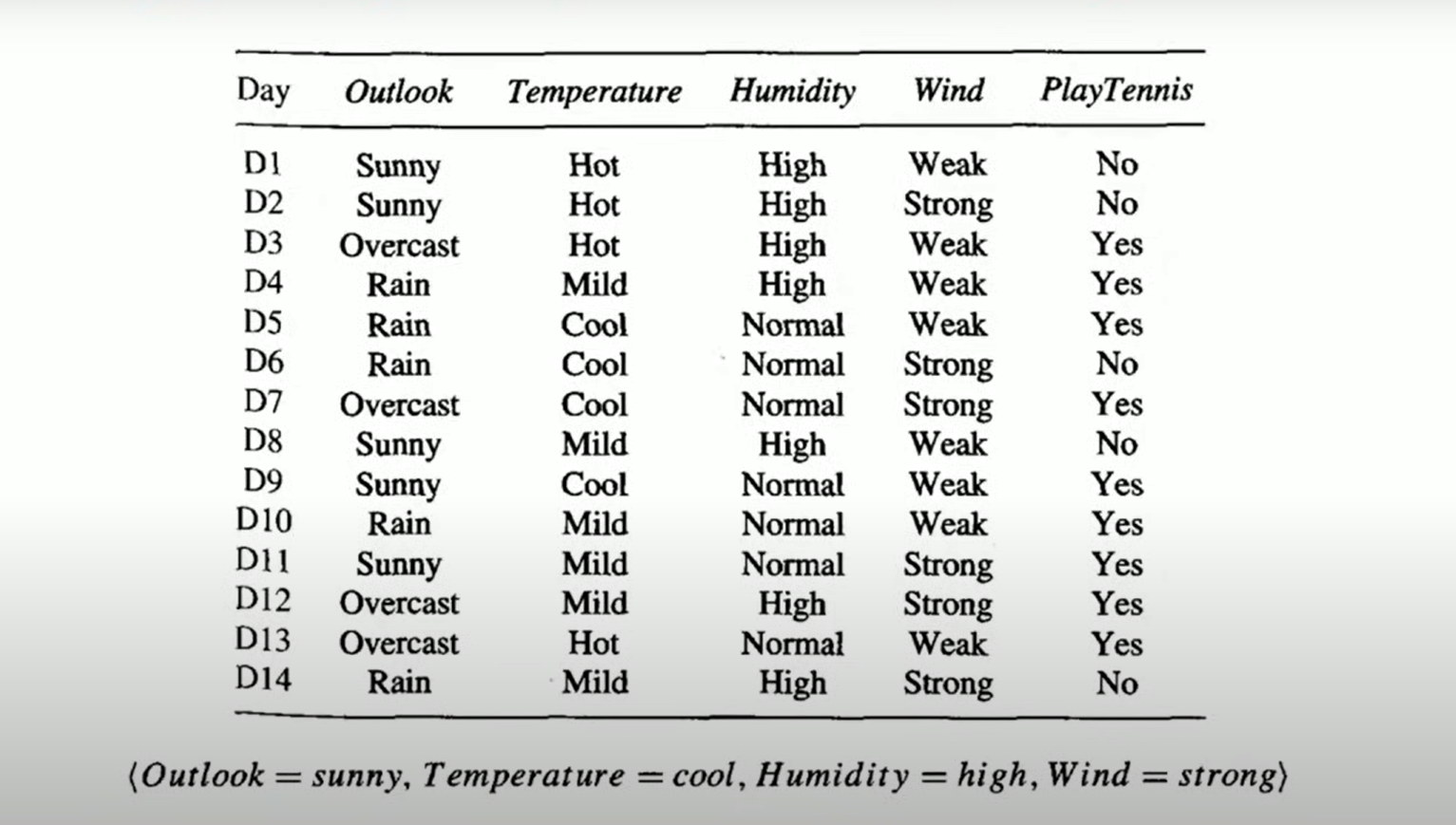
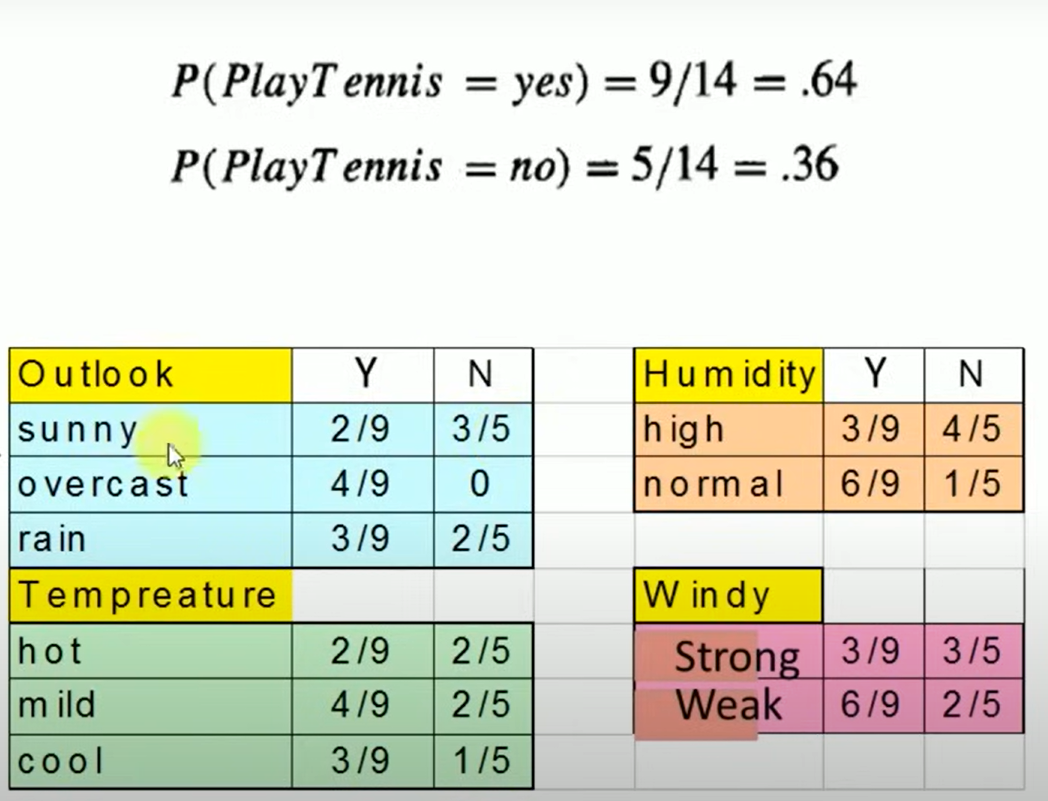
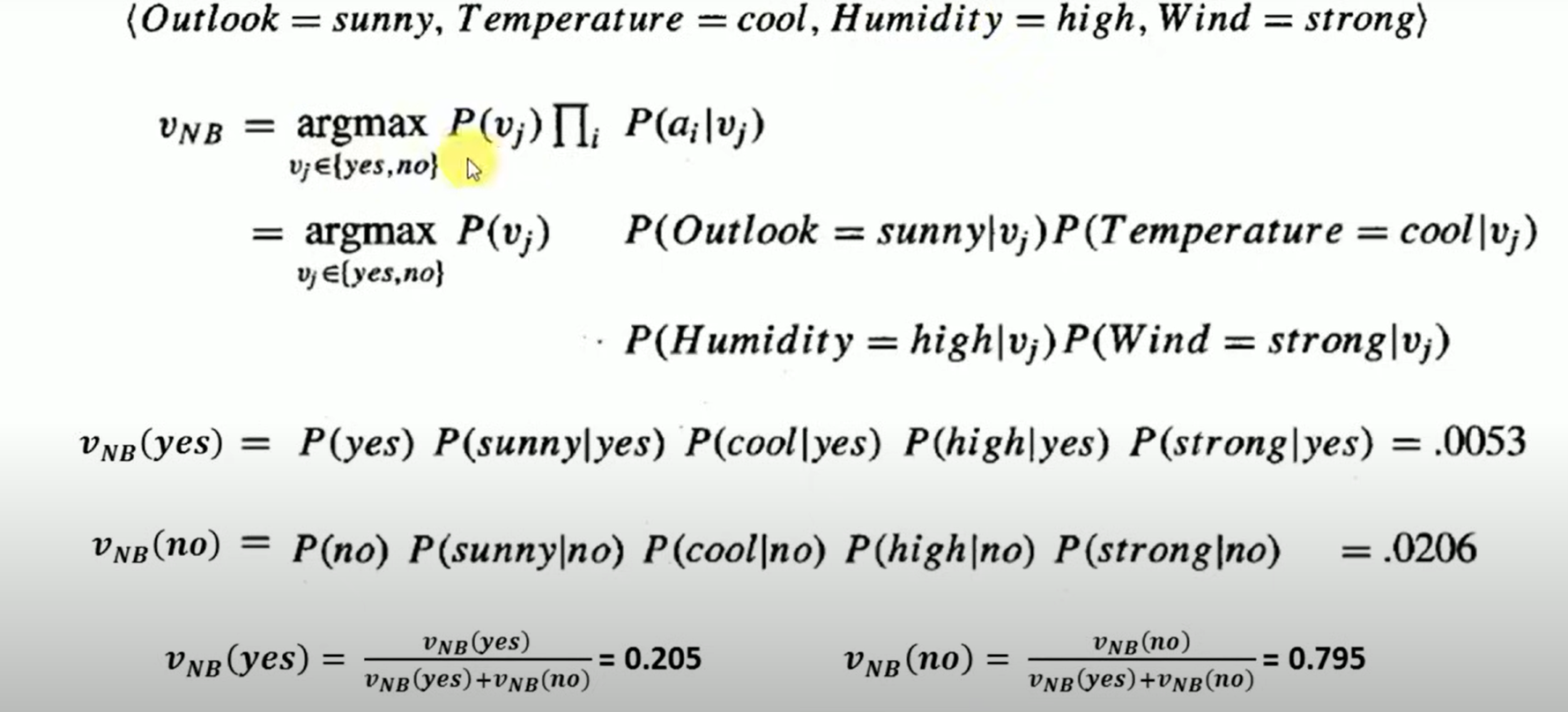
**NAIVEBAYES CLASSIFIER**

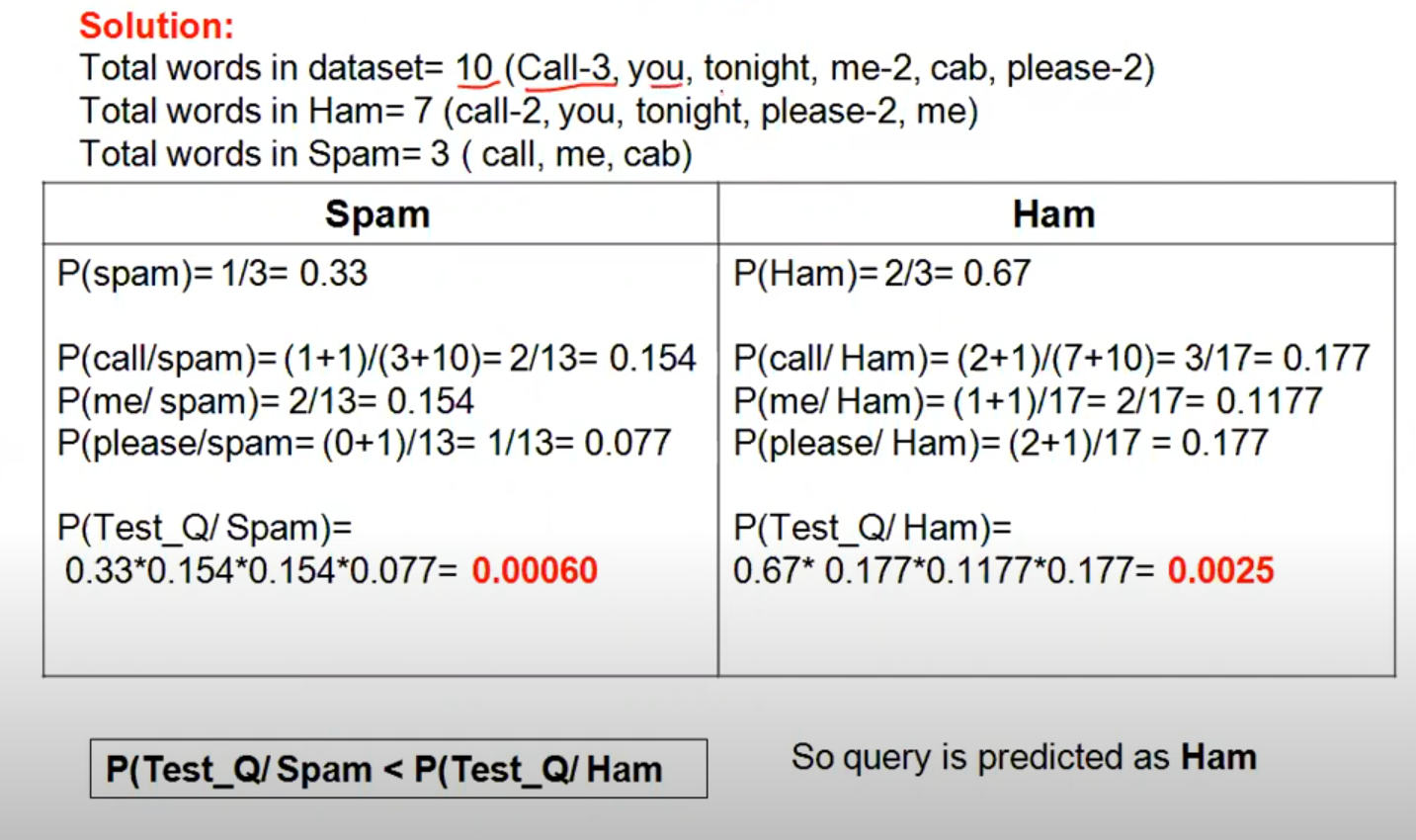




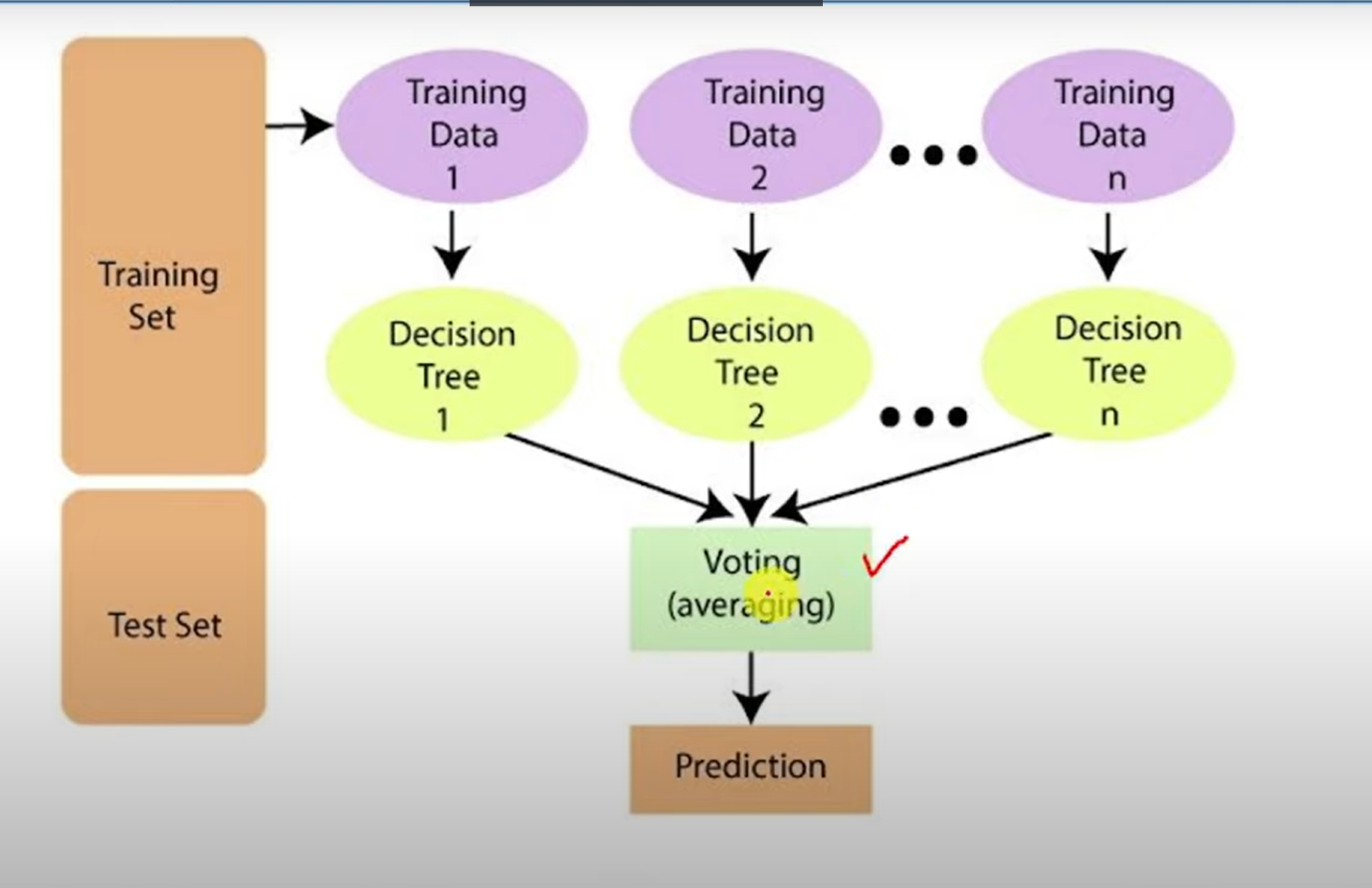


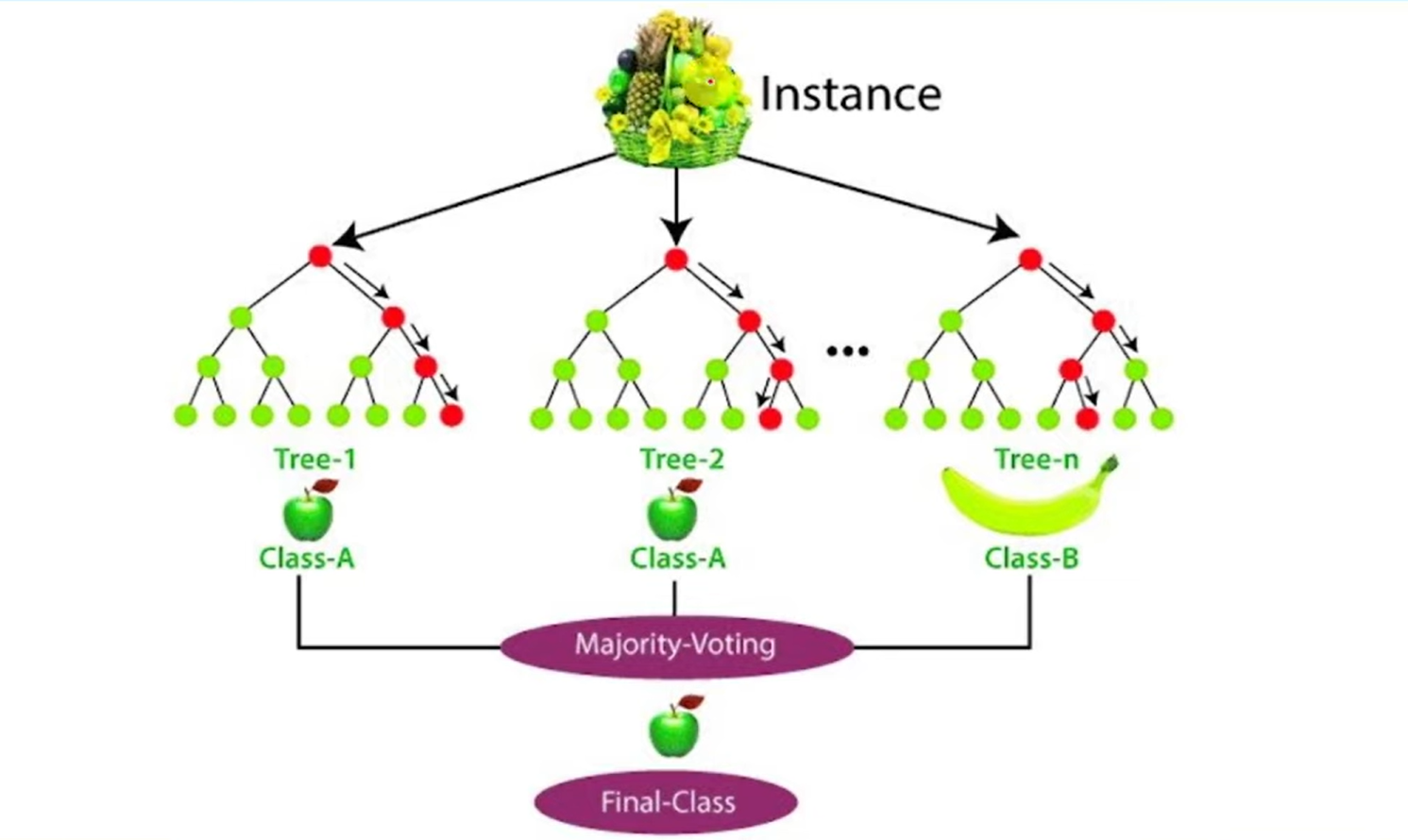
**NAIVEBAYES MULTINOMINAL CLASSIFIER**



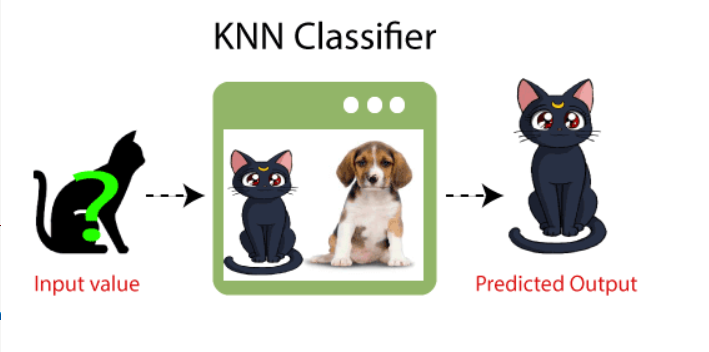


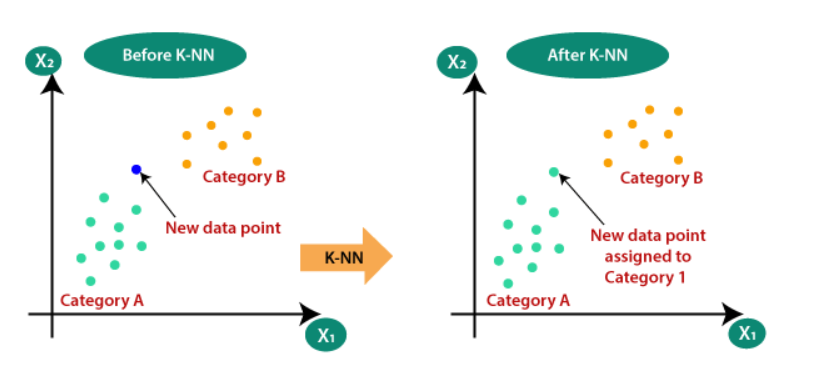
**RANDOM FOREST CLASSIFIER**

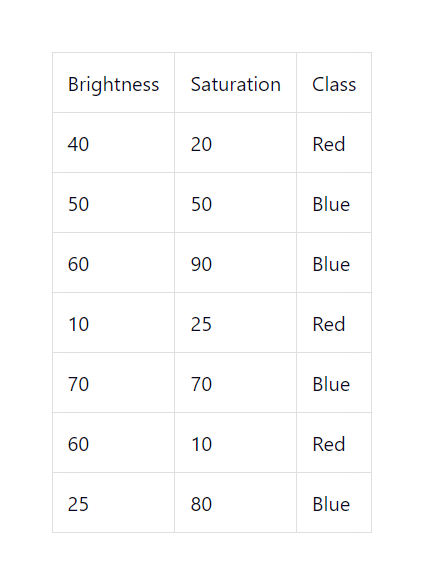
****

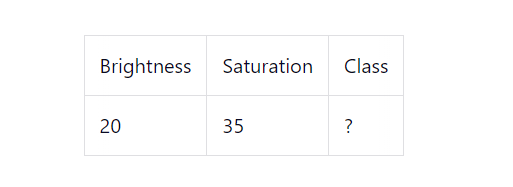
****

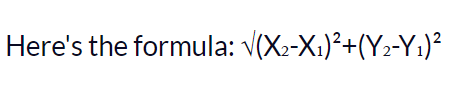
**KNN CLASSIFIER**

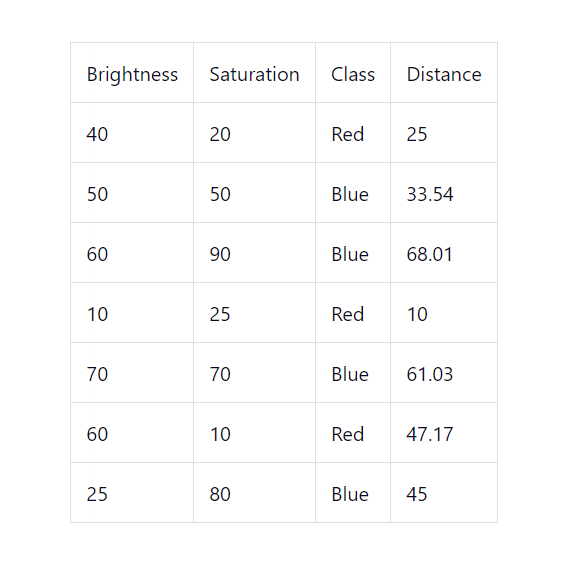
****

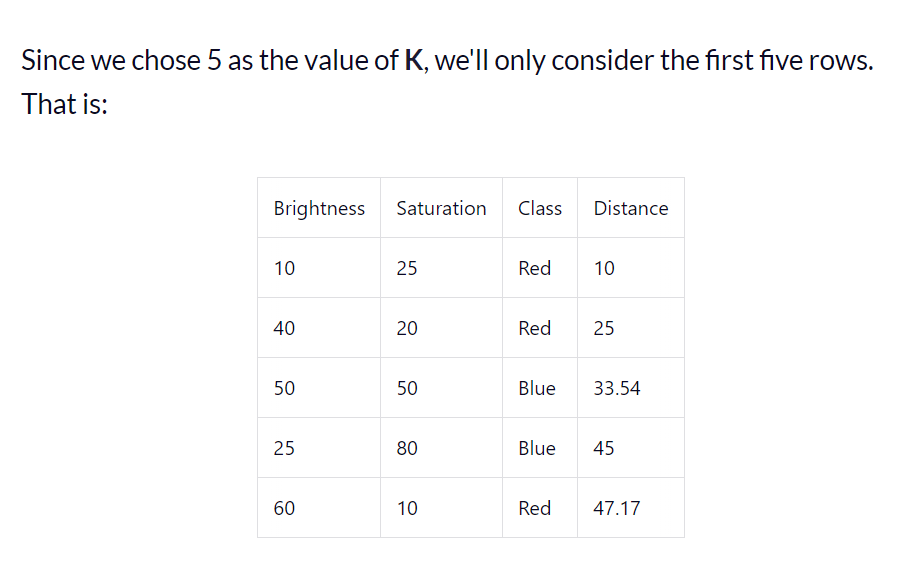
****

****

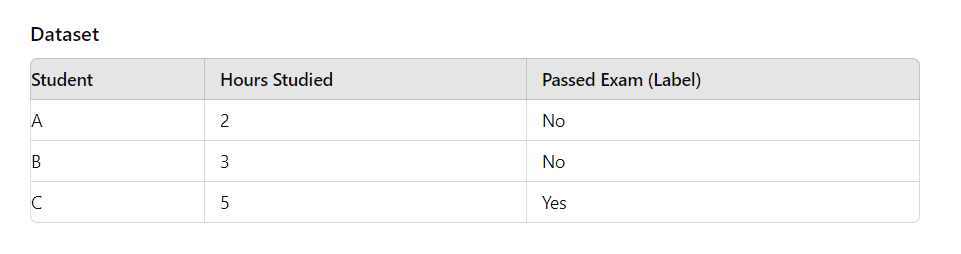
****

****

****

****

**DECISION STUMP CLASSIFIER**

****

Calculate the Gini Impurity for Each Split

The Gini impurity measures how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the set. The formula for Gini impurity for a binary classification problem is:

Gini=1−(p12+p22)\text{Gini} = 1 - \left(p\_1^2 + p\_2^2\right)Gini=1−(p12​+p22​)

Where:

* p1p\_1p1​ is the proportion of class 1 (e.g., "Yes") in a node.
* p2p\_2p2​ is the proportion of class 2 (e.g., "No") in a node.

Split 1: Hours Studied < 2.5

* Left Node (Hours Studied < 2.5): Student A.
  + Passed Exam: No (0 Yes, 1 No)
  + Gini impurity for Left Node:

Ginileft=1−(02+12)=1−1=0\text{Gini}\_{\text{left}} = 1 - \left(0^2 + 1^2\right) = 1 - 1 = 0Ginileft​=1−(02+12)=1−1=0

* Right Node (Hours Studied ≥ 2.5): Students B and C.
  + Passed Exam: No, Yes (1 Yes, 1 No)
  + Gini impurity for Right Node:

Giniright=1−(0.52+0.52)=1−(0.25+0.25)=1−0.5=0.5\text{Gini}\_{\text{right}} = 1 - \left(0.5^2 + 0.5^2\right) = 1 - \left(0.25 + 0.25\right) = 1 - 0.5 = 0.5Giniright​=1−(0.52+0.52)=1−(0.25+0.25)=1−0.5=0.5

* Weighted Gini Impurity of the split:

Weighted Gini=13×0+23×0.5=0+0.333=0.333\text{Weighted Gini} = \frac{1}{3} \times 0 + \frac{2}{3} \times 0.5 = 0 + 0.333 = 0.333Weighted Gini=31​×0+32​×0.5=0+0.333=0.333

Split 2: Hours Studied < 4

* Left Node (Hours Studied < 4): Students A and B.
  + Passed Exam: No, No (0 Yes, 2 No)
  + Gini impurity for Left Node:

Ginileft=1−(02+12)=1−1=0\text{Gini}\_{\text{left}} = 1 - \left(0^2 + 1^2\right) = 1 - 1 = 0Ginileft​=1−(02+12)=1−1=0

* Right Node (Hours Studied ≥ 4): Student C.
  + Passed Exam: Yes (1 Yes, 0 No)
  + Gini impurity for Right Node:

Giniright=1−(12+02)=1−1=0\text{Gini}\_{\text{right}} = 1 - \left(1^2 + 0^2\right) = 1 - 1 = 0Giniright​=1−(12+02)=1−1=0

* Weighted Gini Impurity of the split:

Weighted Gini=23×0+13×0=0\text{Weighted Gini} = \frac{2}{3} \times 0 + \frac{1}{3} \times 0 = 0Weighted Gini=32​×0+31​×0=0

Split 3: Hours Studied < 5

* Left Node (Hours Studied < 5): Students A and B.
  + Passed Exam: No, No (0 Yes, 2 No)
  + Gini impurity for Left Node:

Ginileft=1−(02+12)=1−1=0\text{Gini}\_{\text{left}} = 1 - \left(0^2 + 1^2\right) = 1 - 1 = 0Ginileft​=1−(02+12)=1−1=0

* Right Node (Hours Studied ≥ 5): Student C.
  + Passed Exam: Yes (1 Yes, 0 No)
  + Gini impurity for Right Node:

Giniright=1−(12+02)=1−1=0\text{Gini}\_{\text{right}} = 1 - \left(1^2 + 0^2\right) = 1 - 1 = 0Giniright​=1−(12+02)=1−1=0

* Weighted Gini Impurity of the split:

Weighted Gini=23×0+13×0=0\text{Weighted Gini} = \frac{2}{3} \times 0 + \frac{1}{3} \times 0 = 0Weighted Gini=32​×0+31​×0=0

Step 3: Choose the Best Split

* Split 1 (Hours Studied < 2.5) has a weighted Gini impurity of 0.333.
* Split 2 (Hours Studied < 4) and Split 3 (Hours Studied < 5) both have a weighted Gini impurity of 0.

Since lower Gini impurity is better (indicating a purer split), we choose Split 2 (Hours Studied < 4) as our decision stump, as it perfectly separates the data into pure nodes.

Final Decision Stump

The decision stump will look like this:

* If Hours Studied < 4, predict No (Did not pass).
* Otherwise, predict Yes (Passed).

This stump correctly classifies all the students in our dataset:

* Student A (2 hours): Predicted as "No" (Correct).
* Student B (3 hours): Predicted as "No" (Correct).
* Student C (5 hours): Predicted as "Yes" (Correct).