Decision Tree. (046)

① A story of the Application of Decision Tree.

② Gini-index Model. (CART)
(impurity index / impurity reduction)

③ Entropy Model.
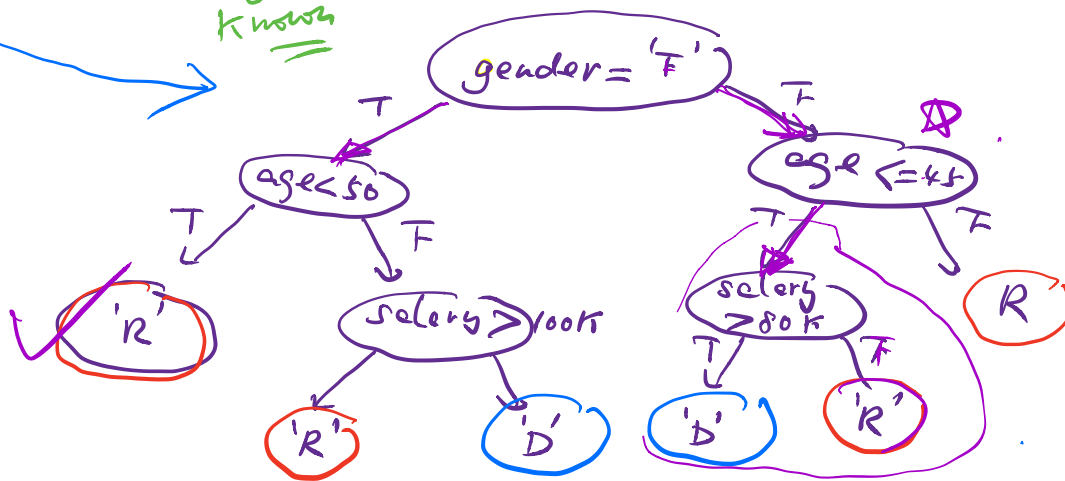(Entropy / Information Gain).

predictor / feature   target / output

No.	age	gender	Salary	Party
1	17	F	60K	R
2	20	M	100K	D
3	40	M	150K	D
4	55	F	160K	D
5	70	M	30K	R
6	30	F	70K	D
7	25	M	110K	D
8	49	M	40K	R
9	60	F	80K	R
10	45	F	100K	? 'R'

training

known

gender = 'F'

T → age < 50
F → age <= 45

age < 50 :
T → 'R'
F → salary > 100K
   T → 'R'
   F → 'D'

age <= 45 :
T → salary > 80K
   T → 'D'
   F → 'R'
F → R

How to Build up a Decision Tree?
→ CART (<u>Gini-index model</u>)
<u>impurity-index</u>.
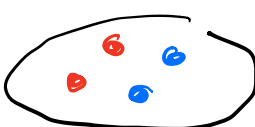
→ How to classify Data
$$1 - \sum_{i=1}^{M} P_i^2$$
\# of class

⟹ To make the "<u>pure</u>"

gini-index =
$$1 - \left(\frac{100\%}{}\right)^2 - (0)^2$$
$$= 0.$$

Data Set - <u>Kristi</u>:
(most pure)
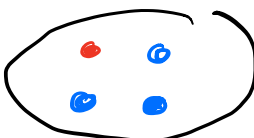
$$1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$
$$= 1 - \frac{1}{4} - \frac{1}{4}$$
$$= \boxed{0.5}$$

Rebecca:
(most impure)

Yogi:
(Mid)

gini-index
$$= 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2$$
$$= 1 - \frac{10}{16} = \frac{6}{16} = \frac{3}{8}$$
$$= \boxed{0.375}$$

Order 3 data set According
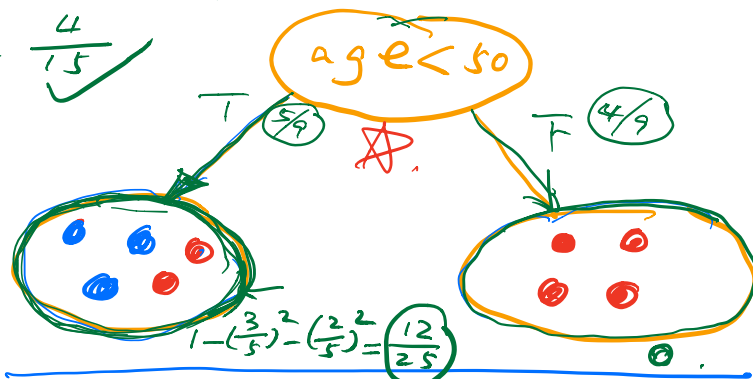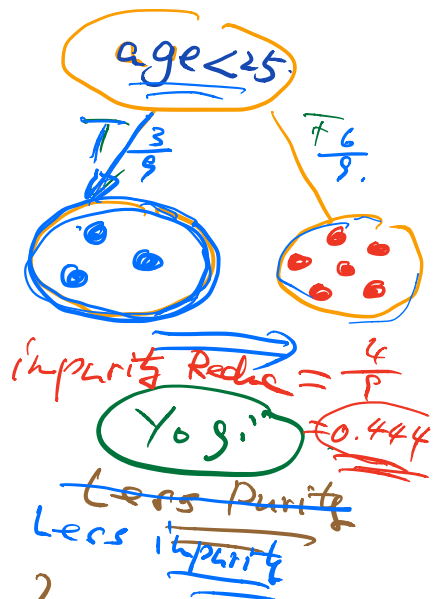to impurity:

pure: homogeneous.'

According to impurity,

Rebecca > Yogi > Kristi.
0.5   0.375   0.

$$gini(age) = \frac{5}{9} \times \frac{12}{25} + \frac{4}{9} \times 0$$

$$= \frac{4}{15} \checkmark$$

age < 50

T ⑤⁄₉   ✗   T ④⁄₉

$1 - (\frac{3}{5})^2 - (\frac{2}{5})^2 = \boxed{\frac{12}{25}}$ 0

Karen

impurity Reduct = 0.67

age < 25 ✓

T $\frac{3}{9}$ T $\frac{6}{9}$

impurity Reduc = $\frac{4}{9}$

Yogi = 0.444

Less Purity
Less Impurity

Data set : { 3 D   6 R }

$$gini\_original\ index = 1 - (\frac{1}{3})^2 - (\frac{2}{3})^2$$

$$= 1 - \frac{1}{9} - \frac{4}{9} = \boxed{\frac{4}{9}} \checkmark$$

$$gini\_by\_Yogi = \frac{3}{9} \cdot \underbrace{gini(age < 25)}_{\downarrow 0}$$

$$+ \frac{6}{9} \cdot \underbrace{gini(age \geq 25)}_{\downarrow 0}$$

$$= 0.$$

$$gini\_reduct = \frac{4}{9} - 0 = \boxed{\frac{4}{9}}$$

For Decision Tree frame

Gini index reduce

$$= \frac{4}{9} - \boxed{\frac{4}{15}} = \underline{\underline{0.26}}.$$

$$= 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$$

$$= 1 - \frac{1}{9} - \frac{4}{9} = \frac{4}{9}.$$

decision fit (_____)