# ML - Assignment

Name: Ch. Keerthana
Reg no: 22BCE9635
slot: G1

1) Construct Decision tree using ID3 Algorithm.

| age | income | student | credit rating | Buys Computer |
|---|---|---|---|---|
| <=30 | high | no | fair | no |
| <=30 | high | no | excellent | no |
| 31-40 | high | no | fair | yes |
| >40 | medium | no | fair | yes |
| >40 | low | yes | fair | yes |
| >40 | low | yes | excellent | no |
| 31-40 | low | yes | excellent | yes |
| <=30 | medium | no | fair | no |
| <=30 | low | yes | fair | yes |
| >40 | medium | yes | fair | yes |
| <=30 | medium | yes | excellent | yes |
| 31-40 | medium | no | excellent | yes |
| 31-40 | high | yes | fair | yes |
| >40 | medium | no | excellent | no |

A Attribute : Age

$$\text{Entropy}(S) = -\frac{9}{14}\log_2(9/14) - \frac{5}{14}\log_2(5/14)$$

"yes" probability, "NO" probability (for Buys Computer)

$$\text{Entropy}(S) = -\frac{9}{14}\log_2(9/14) - \frac{5}{14}\log_2(5/14) = 0.94$$

- age $\leq 30$ ( 2 yes and 3 no)
- age 31-40 ( 4 yes and 0 no)
- age $>40$ ( 3 yes & 2 no)

Entropy (Age):

(i) $\leq 30$

$$\text{Entropy} = \frac{5}{14}\left[\frac{-2}{5}\log_2(2/5) - \frac{3}{5}\log_2(3/5)\right]$$

$$= \frac{5}{14}(0.9709)$$

(ii) age: 31-40

Entropy $= \frac{4}{14}(0)$ → because, for age 31-40 only 'yes' was there so, Entropy value will be '0'.

(iii) age: $>40$

$$\text{Entropy} = \frac{5}{14}\left[\frac{-3}{5}\log_2(3/5) - \frac{2}{5}\log_2(2/5)\right]$$

$$= \frac{5}{14}(0.9709)$$

$$\text{Entropy (age)} = \frac{5}{14}(0.9709) + 0 + \frac{5}{14}(0.9709)$$

$$= 0.6935$$

$$\therefore \text{Gain (age)} = \underset{\text{total}}{\text{Entropy (S)}} - \text{Entropy (age)}$$

$$= 0.94 - 0.6935 = 0.2465$$

Attribute: Income

Entropy (S) = 0.94
(whole data set)

- Income (high) → 2 yes & 2 no
- Income (medium) → 4 yes & 2 no
- Income (low) → 3 yes & 1 no

$$\text{Entropy (Income)} = \frac{4}{14}\left[-\frac{2}{4}\log_2(2/4) - \frac{2}{4}\log_2(2/4)\right] \text{ (high)}$$

$$+ \frac{6}{14}\left[-\frac{4}{6}\log_2(4/6) - \frac{2}{6}\log_2(2/6)\right] + \frac{4}{14}\left[-\frac{3}{4}\log_2(3/4) - \frac{1}{4}\log_2(1/4)\right]$$

↳ (medium)

$$= \frac{4}{14}(1) + \frac{6}{14}(0.918) + \frac{4}{14}(0.811)$$

$$= 0.285714 + 0.393428 + 0.231714 = 0.9108$$

$$\text{Gain (Income)} = \text{Entropy}(S) - \text{Entropy (Income)}$$

$$= 0.94 - 0.9108 = 0.0292$$

Attribute : Student

- Student (yes) ⇒ (6 yes & 1 NO)
- Student (NO) ⇒ (3 yes & 4 NO)

$$\text{Entropy (Student)} = \frac{7}{14}\left[-\frac{6}{7}\log_2(6/7) - \frac{1}{7}\log_2(1/7)\right] +$$

$$\frac{7}{14}\left[-\frac{3}{7}\log_2(3/7) - \frac{4}{7}\log_2(4/7)\right]$$

$$= \frac{7}{14}(0.5916) + \frac{7}{14}(0.9852)$$

$$= 0.2958 + 0.4926 = 0.7884$$

$$\therefore \text{Gain (Student)} = 0.94 - 0.7884 = 0.1516.$$

Attribute : Credit_Rating

- Credit_rating (fair) ⇒ (6 yes & 2 NO)
- Credit_rating (excellent) ⇒ (3 yes & 3 NO)

$$\text{Entropy (Credit\_rating)} = \frac{8}{14}\left[-\frac{6}{8}\log_2(6/8) - \frac{2}{8}\log_2(2/8)\right] + \frac{6}{14}\left[-\frac{3}{6}\log_2(3/6) - \frac{3}{6}\log_2(3/6)\right]$$

$$= \frac{8}{14}(0.8112) + \frac{6}{14}(1) = 0.4635 + 0.4285$$

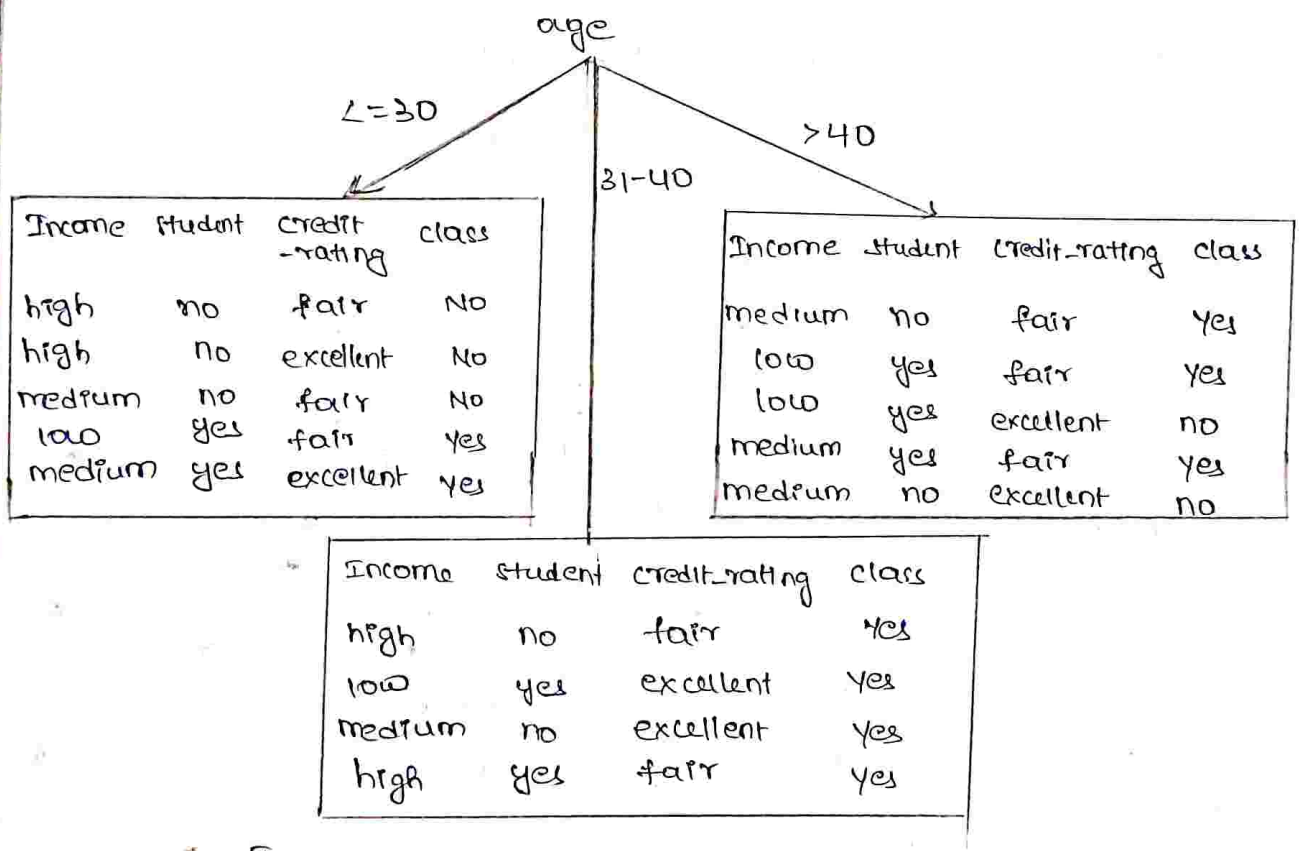
$$= 0.8920$$

$$\text{Gain(Credit\_rating)} = 0.94 - 0.8920$$

$$= 0.048$$

By Comparing Information Gain for Age, Student, Income and Credit_rating, Age is having more Gain so, we consider Age as root node and start constructing the tree.

=> under Age (root node) we have 3 more conditions i.e, <=30, 31-40 & >40 so, we construct subsets for those.




| Income | Student | Credit_rating | Class |
|---|---|---|---|
| high | no | fair | No |
| high | no | excellent | No |
| medium | no | fair | No |
| low | yes | fair | Yes |
| medium | yes | excellent | Yes |

| Income | Student | Credit_rating | Class |
|---|---|---|---|
| medium | no | fair | Yes |
| low | yes | fair | Yes |
| low | yes | excellent | no |
| medium | yes | fair | Yes |
| medium | no | excellent | no |

| Income | Student | Credit_rating | Class |
|---|---|---|---|
| high | no | fair | Yes |
| low | yes | excellent | Yes |
| medium | no | excellent | Yes |
| high | yes | fair | Yes |

Attribute: Income

$E(S_{age<=30}) = E(2,3) = -\frac{2}{5}\log_2(2/5) - \frac{3}{5}\log_2(3/5) = 0.97$

Entropy(Income) $= \frac{2}{5}(0) + \frac{2}{5}(1) + \frac{1}{5}(0) = \frac{2}{5}(1) = 0.4$

Gain(Income) $= 0.97 - 0.4 = 0.57$

Attribute: (Student)

$E(S_{age<=30}) = E(2,3) = -\frac{2}{5}\log_2(2/5) - \frac{3}{5}\log_2(2/5) = 0.97$

Entropy(Student) $= 2/5(0) + 3/5(0) = 0$

Gain(student) $= (0.97 - 0) = 0.97$.

Attribute (credit_rating)

$$\text{Entropy (credit)} = \frac{3}{5}\left(-\frac{1}{3}\log_2(1/3) - \frac{2}{3}\log_2(2/3)\right) + \frac{2}{5}(1)$$

$$= 0.9508$$

$$\text{Gain(Credit}_{\text{rating}}) = 0.97 - 0.9508 = 0.0192$$

for Age >40)

$$E(S_{age>40}) = E(3,2) = -\frac{3}{5}\log_2(3/5) - \frac{2}{5}\log_2(2/5) = 0.97$$

$$\text{Entropy (Income)} = \frac{3}{5}\left[-\frac{2}{3}\log_2(2/3) - \frac{1}{3}\log_2(1/3)\right] + \frac{2}{5}(1)$$

$$= \frac{3}{5}(0.9182) + \frac{2}{5} = 0.95$$

Gain(Income) = (0.97 - 0.95) = 0.02

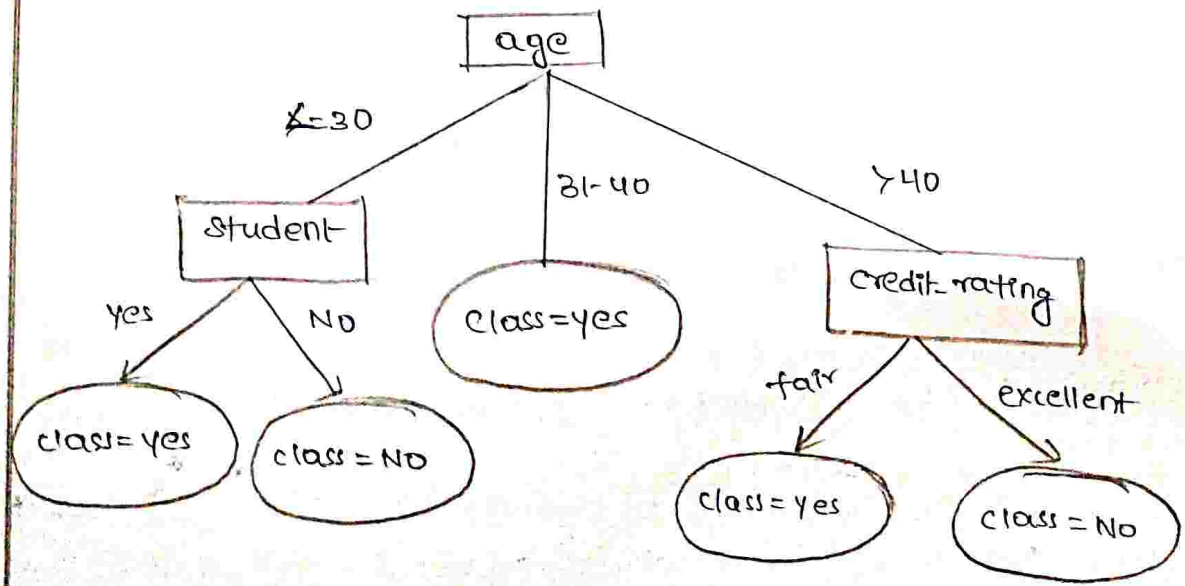
$$\text{Entropy (student)} = \frac{3}{5}\left[-\frac{2}{3}\log_2(2/3) - \frac{1}{3}\log_2(1/3)\right] + \frac{2}{5}(1) = 0.95$$

Gain (student) = (0.97 - 0.95) = 0.02

$$\text{Entropy (credit}_{\text{rating}}) = \frac{3}{5}(0) + \frac{2}{5}(0) = 0$$

Gain (Credit rating) = (0.97 - 0) = 0.97

Final decision tree:



2) Construct Decision tree using CART algorithm

| CGPA | Inter active | practical knowledge | Common skills | Job offer |
|---|---|---|---|---|
| >=9 | Yes | very good | Good | Yes |
| >=8 | No | Good | Moderate | Yes |
| >=9 | No | Average | poor | No |
| <8 | No | Average | Good | No |
| >=8 | Yes | Good | Moderate | Yes |
| >=9 | Yes | Good | Moderate | Yes |
| <8 | Yes | Good | poor | No |
| >=9 | No | Very good | Good | Yes |
| >=8 | Yes | Good | Good | Yes |
| >=8 | Yes | Average | Good | Yes |

Step-1: calculate the Gini_index for dataset.

$$\text{Gini-index}(T) = 1 - \sum_{i=1}^{2} P_i^2$$

Gini_Index (T) = $1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2 = 1 - 0.49 - 0.09 = 1 - 0.58$
(Jobotter)

$= 0.42$ (whole dataset (T))

Step-2: compute Gini-Index for each attribute and each of the subset in the attribute.

i) CGPA: (>=9, >=8, <8)

possible subsets

↳ {}, {>=9}, {>=8}, {>=9,>=8}, {>=9,>8}, {>=8,>8}, {>=9,>=8,>8}

calculation of best splitting subset:

$$\text{Gini-Index}(T,A) = \left|\frac{S_1}{T}\right| \text{Gini}(S_1) + \left|\frac{S_2}{T}\right| \text{Gini}(S_2)$$

(i)

Gini-Index(T, CGPA∈{>=9,>8}) = $1 - (7/8)^2 - (1/8)^2 = 1 - 0.7806 = 0.2194$

Gini_Index (T, CGPA∈{<8}) = $1 - 0(2)^2 - 2(2)^2 = 1 - 1 = 0$

Gini_Index (T, CGPA ∈ {{>=9, >=8}, {<8}}) = (8/10) (0.2194) + (2/10)(0)
= 0.17552

(ii)

Gini-Index (T, CGPA ∈ {≥9, <8}) = 1-(3/6)²-(3/6)² = (1-0.5) = 0.5

Gini-Index (T, CGPA ∈ {≥8}) = 1-(4/4)²- (0/4)² = (1-1) = 0

Gini-Index ((T, CGPA ∈ {{≥9, ≥8}, {≥9, <8}}) = (6/10) 0.5 + (4/10) 0 = 0.3

(iii)

Gini-Index (T, CGPA ∈ {≥8, <8}) = 1-(4/6)²-(2/6)² = 1-0.555 = 0.445

Gini-Index (T, CGPA ∈ {≥9}) = 1-(3/4)²-(1/4)² = 1-0.625 = 0.375

Gini_Index (T, CGPA ∈ {{≥8, <8}, {≥9}}) = (6/10) 0.445 + (4/10) 0.375
= 0.417

Step-3:

In the above 3 cases 1st case subset is having minimum Gini_Index value, so, 1st subset is the best splitting subset

→ the subset CGPA = {{>=9, >=8}, <8} has low T value

Step-4:

Compute (ΔGini / best splitting subset) of that attribute.

ΔGini (CGPA) = Gini(T) - Gini (T, CGPA)
= (0.42-0.1755) = 0.2445

Repeat the same process for remaining attributes in dataset such as for interactiveness, practical knowledge and Communication skills.

Categories for interactiveness:

Gini-Index (T, interactiveness ∈ {yes}) = 1-(5/6)²-(1/6)² = (1-0.72) = 0.28

Gini-Index (T, intr ∈ {No}) = 1-(2/4)²-(2/4)² = (1-0.5) = 0.5

Gini-Index (T, intr ∈ {yes, No}) = 6/10 (0.28) + 4/10 (0.5)
= 0.168 + 0.2 = 0.368

ΔGini (Interactiveness) = Gini(T) - Gini(T, Int) = (0.42-0.368) = 0.052

Categories for practical knowledge

(i) Gini-Indx(T, pract know ∈ {very good, Good}) = $1-(5/7)^2-(1/7)^2$

$= 0.2456$

Gini-Index(T, pract know ∈ {Avg}) = $1-(1/3)^2-(2/3)^2 = 1-0.555 = 0.445$

Gini-Index(T, pract know ∈ {{VG, G}, Avg}) = $(7/10)\,0.2456 + (3/10)\,0.445$

$= 0.3054$

(ii) G-I(T, pract know ∈ {VG, Avg}) = $1-(3/5)^2-(2/5)^2 = 1-0.52 = 0.48$

G-I(T, pract know ∈ {Good}) = $1-(4/5)^2-(1/5)^2 = 1-068 = 0.32$

G-I(T, pract know ∈ {{VG, Avg}, G}) = $5/10\,(0.48) + 5/10\,(0.32) = 0.40$

(iii)

G-I(T, pract know ∈ {G, Avg}) = $1-(5/8)^2-(2/8)^2 = 1-0.5312 = 0.4688$

G-I(T, pract know ∈ {VG}) = $1-(2/2)^2-(0/2)^2 = (1-1) = 0$

G-I(T, pract know ∈ {{G, Avg}, VG}) = $(8/10)\,0.4688 + (2/10)\,0 = 0.3750$

case-(i) is having low Gini-index value so, 1st subset is best splitting subset

$\Delta$Gini(pract-know) = Gini(T) − Gini(T, pract know)

$= 0.42 - 0.3054 = 0.1146$

Categories for ~~Communication~~ Common skills

↳ comsk

(i) G-I(T, comsk ∈ {G, Mod}) = $1-(7/8)^2-(1/8)^2 = 1-0.7606 = 0.2194$

G-I(T, comsk ∈ {poor}) = $1-(2/2)^2-(0/2)^2 = (1-1) = 0$

G-I(T, comsk ∈ {{G, Mod}, P}) = $(8/10)\,0.2194 + (2/10)\,0 = 0.1755$

(ii) G-I(T, comsk ∈ {G, P}) = $1-(4/7)^2-(3/7)^2 = 1-0.5101 = 0.4899$

G-I(T, comsk ∈ {Mod}) = $1-(3/3)^2-(0/3)^2 = (1-1) = 0$

G-I(T, comsk ∈ {{G, P}, Mod}) = $(7/10)\,0.4899 + (3/10)\,0$

$= 0.3429$

(iii)

$G\text{-}I(T, \text{comsk} \in \{\text{Hod}, P\}) = 1-(3/5)^2-(2/5)^2 = 1-0.52 = 0.48$

$G\text{-}I(T, \text{comsk} \in \{G\}) = 1-(4/5)^2-(1/5)^2 = 1-0.68 = 0.32$

$G\text{-}I(T, \text{comsk} \in \{\{\text{Hod}, P\}, G\}) = (5/10)^2 0.48 + (5/10)^2 0.32 = 0.40$

$\Delta\text{Gini}(\text{comsk}) = \text{Gini}(T) - \text{Gini}(T, \text{comsk})$ (1st subset → best splitting subset)

$= 0.42 - 0.1955 = 0.2445$

Step-5:

After calculating Gini-index and ΔGini for all attributes we need to consider highest ΔGini value attribute as root node here, for CGPA & comsk having highest ΔGini value. so, let us Consider CGPA as root node

⇒ for subset $\{\geq 9, \geq 8\}$ have both yes and No values so, for deleting that again calculate Gini-Index value

$\text{Gini-Index}(T) = 1-(7/8)^2-(1/8)^2 = 1-0.766-0.0156 = 0.2184$

(i) Interactiveness

$G\text{-}I(T, \text{intr} \in \{\text{Yes}\}) = 1-(5/5)^2-(0/5)^2 = 0$

$G\text{-}I(T, \text{intr} \in \{\text{No}\}) = 1-(2/3)^2-(1/3)^2 = 1-0.44-0.111 = 0.449$

$G\text{-}I(T, \text{intr} \in \{\text{Yes}, \text{No}\}) = 5/8(0) + 1/8(0.449) = 0.056$

$\Delta\text{Gini}(\text{Intr}) = \text{Gini}(T) - \text{Gini}(T, \text{intr})$

$= (0.2184 - 0.056) = 0.1624$

(ii) practical knowledge

$G\text{-}I(T, \text{intr}\{\{VG, G\}, G\}) = 0.125$

$\Delta\text{Gini}(\text{Pract Know}) = \text{Gini}(T) - \text{Gini}(T, \text{Pract Know})$

$= 0.2184 - 0.125$

$= 0.0934$

(iii) Communication skills

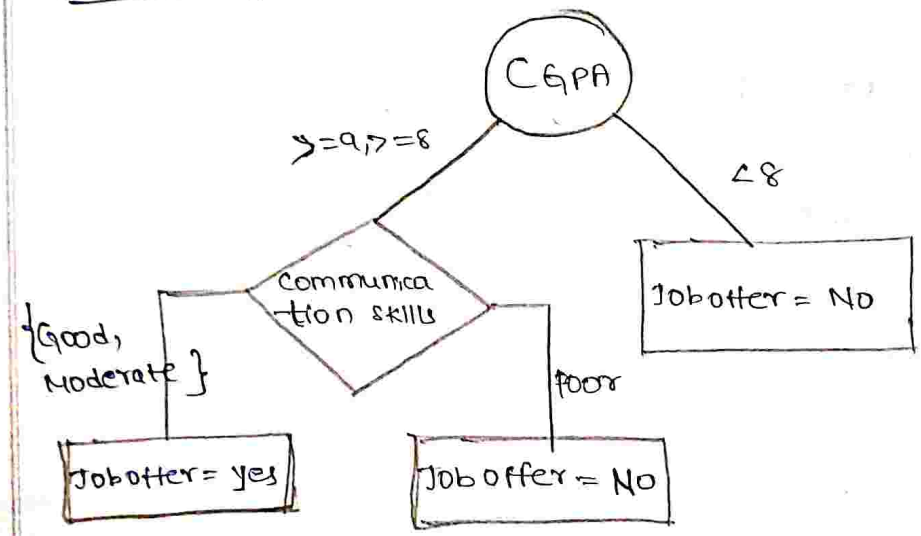
| subsets | Gini- Index |
|---|---|
| {{G, Mod}, P} | 0 |
| {{G, P}, Mod} | 0.2 |
| {{Mod, P}, G} | 0.1875 |

$\Delta Gini(Comsk) = Gini(T) - Gini(T, comsk)$

$= 0.2184 - 0 = 0.2184$

∴ Communication skills having highest $\Delta Gini$ value (high priority)

Decision tree:



3) Construct Decision tree using C4.5 Algorithm

| CGPA | Interactive | Practical Knowledge | Communication Skills | Job offer |
|---|---|---|---|---|
| >=9 | yes | very good | Good | yes |
| >=8 | No | Good | Moderate | yes |
| >=9 | No | Average | poor | No |
| <8 | No | Average | Good | No |
| >=8 | yes | Good | Moderate | yes |
| >=9 | yes | Good | Moderate | yes |
| <8 | yes | Good | poor | No |
| >=9 | No | very Good | Good | yes |
| >=8 | yes | Good | Good | yes |
| >=8 | yes | Average | Good | yes |

Step-1:

calculate class-Entropy for target class 'Job offer'.

$$\text{Entropy-Info}(T) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

$$\text{Entropy-Info}(\underset{T,3}{\text{Job offer}}) = \left[7/10 \log_2(7/10) + 3/10 \log_2(3/10)\right]$$

$$= (-0.3599 + 0.5208)) = 0.8807$$

Step-2:

calculate the Entropy-info, Gain (Info-gain), Split-Info, Gain-ratio for each attribute in training dataset.

$$\text{Entropy-Info}(T, A) = \sum_{i=1}^{n} \left|\frac{A_i}{T}\right| \times \text{Entropy-Info}(A_i)$$

(1) CGPA:

$$\text{Entropy-Info}(T, \text{CGPA}) = \frac{4}{10}\left[-3/4 \log_2(3/4) - 1/4 \log_2(1/4)\right] + \frac{4}{10}\left[-4/4 \log_2(4/4)\right]$$

$$+ \frac{2}{10}\left[-0/2 \log_2(0/2) - \frac{2}{2} \log_2(2/2)\right] = \frac{4}{10}\left[0.3111 + 0.4997\right] + 0 + 0$$

$$= 0.3243$$

Info_Gain(A) = Entropy-Info(T) − Entropy-info(T,A)

Gain(CGPA) = (0.8807 − 0.3243) = 0.5564

$$\text{split\_info}(T,A) = -\sum_{i=1}^{K} \frac{|A_i|}{|T|} \times \log_2\left(\frac{|A_i|}{|T|}\right)$$

split_info(CGPA) = $-\frac{4}{10}\log_2(4/10) - \frac{4}{10}\log_2(4/10) - \frac{2}{10}\log_2(2/10)$

= (0.5285 + 0.5285 + 0.4641) = 1.5211

Gain ratio(CGPA) = $\frac{0.5564}{1.5211}$ = 0.3658

(2) Interactiveness:

Entropy-info(T, intr) = $\frac{6}{10}\left[-5/6\log_2(5/6) - 1/6\log_2(1/6)\right] + \frac{4}{10}\left[-2/4\log_2(2/4) - \frac{2}{4}\log_2(2/4)\right]$

interactive

= $\frac{6}{10}[0.2191 + 0.4306] + \frac{4}{10}[0.4997 + 0.4997]$

= (0.3898 + 0.3998) = 0.7896

Gain(Intr) = (0.8807 − 0.7896) = 0.0911

split-info(T, intr) = $-\frac{6}{10}\log_2(6/10) - \frac{4}{10}\log_2(4/10)$ = 0.9704

Gain ratio(intr) = $\frac{0.0911}{0.9704}$ = 0.0939

(3) Practical Knowledge:

Entropy-info(T, practKn) = $\frac{2}{10}\left[-2/2\log_2(2/2)\right] + \frac{3}{10}\left[-1/3\log_2(1/3) - 2/3\log_2(2/3)\right]$

practical knowledge

$+ \frac{5}{10}\left[-4/5\log_2(4/5) - 1/5\log_2(1/5)\right]$

= 2/10(0) + 3/10(0.5280 + 0.3897) + 5/10(0.2574 + 0.4641)

= 0 + 0.2573 + 0.3608 = 0.6361

Gain(pract Kn) = 0.8807 − 0.6361 = 0.2448

Split info(T, pract Kn) = $-\frac{2}{10}\log_2(2/10) - 5/10\log_2(5/10) - 3/10\log_2(3/10)$

= 1.4853

Gain_ratio(Pract Kno) = $\frac{0.2448}{1.4853}$ = 0.1648

(4) Communication skills

Entropy - info (T, comsk) = $\frac{5}{10}\left[-4/5 \log_2(4/5) - 1/5 \log_2(1/5)\right] +$

$\frac{3}{10}\left[-3/3 \log_{10}(3/3)\right] + \frac{2}{10}\left[-2/2 \log_{10}(2/2)\right]$

$= 5/10\ (0.5280 + 0.3897) + 3/10(0) + 2/10(0)$

$= 0.3609$

Gain(comsk) = $(0.8813 - 0.3609) = 0.5202$

Split - info (T, comsk) = $-5/10 \log_2(5/10) - 3/10 \log_2(3/10) - 2/10 \log_2(2/10)$

common skills $= 1.4853$

Gain_ratio (comsk) = $\left(\frac{0.5202}{1.4853}\right) = 0.3502$

CGPA is root node

for CGPA ($\geq 9$)

Entropy - info(job offer) = $-3/4 \log_2(3/4) - 1/4 \log_2(1/4)$

$= 0.3112 + 0.5 = 0.8112$

Entropy - info(T, intr) = $2/4\left[-2/2 \log_2(2/2)\right] + 2/4\left[2(\frac{-1}{2}) \log_2(1/2)\right] = 0.4997$

Gain(intr) = $0.8112 - 0.4997 = 0.3111$

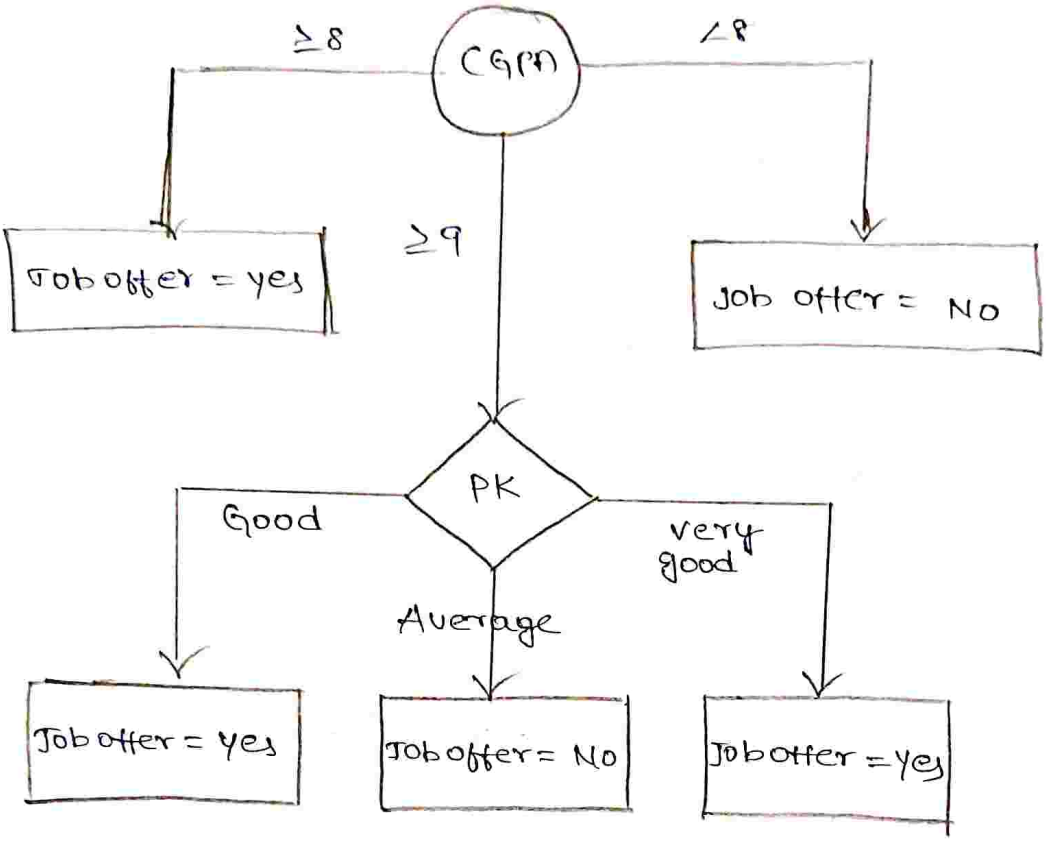
split - info (T, intr) = $-2/4 \log_2(2/4) - 2/4 \log_2(2/4) = (0.5 + 0.5) = 1$

Gain ratio = $\frac{0.3112}{1} = 0.3112$

Gain - ratio

| Attributes | Gain-ratio |
|---|---|
| interactiveness | 0.3112 |
| practical Knowledge | 0.5408 |
| Communication skills | 0.5408 |

Decision tree :



4) Implement KNN classifier (only for classification)

Example:

Given data query => $x = (Maths = 6, CS = 8)$

and $K = 3$ - nearest neighbour

classification - pass/fail

| S.NO | Maths | CS | Result |
|---|---|---|---|
| 1) | 4 | 3 | F |
| 2) | 6 | 7 | P |
| 3) | 7 | 8 | P |
| 4) | 5 | 5 | F |
| 5) | 8 | 8 | P |

Step-1:

Calculate Euclidean distance (d)

$$d = \sqrt{(x_{O_1} - x_{A_1})^2 + (x_{O_2} - x_{A_2})^2}$$

O → observed value
A → Actual value.

$d_1 = \sqrt{(6-4)^2 + (8-3)^2} = \sqrt{4+25} = \sqrt{29} = 5.38$

$d_2 = \sqrt{(6-6)^2 + (8-7)^2} = \sqrt{0+1} = 1$

$d_3 = \sqrt{(6-7)^2 + (8-8)^2} = \sqrt{1+0} = 1$

$d_4 = \sqrt{(6-5)^2 + (8-5)^2} = \sqrt{1+9} = \sqrt{10} = 3.16$

$d_5 = \sqrt{(6-8)^2 + (8-8)^2} = \sqrt{4+0} = \sqrt{4} = 2$

Step-2: The distances that are closer and less than K value are 1, 1, 2 ie, $d_2$, $d_3$ and $d_5$ their, result values are Pass

[∴ classification - pass]

5) Implement Linear discriminant analysis (LDA) for suitable example.

$x_1 = \{ (4,1), (2,4), (2,3), (3,6), (4,4) \}$

$x_2 = \{ (9,10), (6,8), (9,5), (8,7), (10,8) \}$

Step - 1:

Calculate means

$$\mu_1 = \begin{bmatrix} \frac{4+2+2+3+4}{5} \\ \frac{1+4+3+6+4}{5} \end{bmatrix} = \begin{bmatrix} 3 \\ 3.6 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} \frac{9+6+9+8+10}{5} \\ \frac{10+8+5+7+8}{5} \end{bmatrix} = \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}$$

Step - 2:

Scatter Matrix

$$S_i = \sum_{q=1}^{n} (x_i - \text{mean}x_i)(x_i - \text{mean}x_i)^T$$

$$S_1 = \left[ (4-3, 1-3.6)(4-3, 1-3.6)^T + (2-3, 4-3.6)(2-3, 4-3.6)^T + (-1)(-0.6)(2-3)(3-3.6)^T + (3-3)(6-3.6)(6-3.6)^T + (4-3)(4-3.6)^T + (4-3)(4-3.6) \right]$$

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

$$S_2 = \left[ (9-8.4)(10-7.6)(9-8.4)(10-7.6)^T + (6-8.4)(8-7.6)(6-8.4)(8-9.6)^T + (9-8.4)(5-7.6)(9-8.4)(5-7.6)^T + (8-8.4)(7-7.6)(8-8.4)(7-7.6)^T + (10-8.4)^2 (8-7.6)(8-7.6)^T \right]$$

Q-

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_w = (S_1 + S_2)$$

$$S_w = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix} + \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_w = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Step - 3:

Scatter matrix $S_B$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \begin{bmatrix} 3 - 8.4 \\ 3.6 - 7.6 \end{bmatrix} \begin{bmatrix} 3 - 8.4 \\ 3.6 - 7.6 \end{bmatrix}^T$$

$$= \begin{bmatrix} -5.4 \\ -4 \end{bmatrix} \begin{bmatrix} -5.4 & -4 \end{bmatrix}$$

$$= \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}$$

$$S_w^{-1} = \frac{\text{Adj}(S_w)}{|S_w|}$$

$$|S_w| = (2.64)(5.28) - (0.44)^2$$

$$|S_w| = 13.93 - 0.19$$

$$|S_w| = 13.73$$

$$\text{Adj}(S_w) = \begin{bmatrix} 16 & -21.6 \\ -21.6 & 29.16 \end{bmatrix}$$

$$S_\omega^{-1} S_B = \frac{1}{13.73} \begin{bmatrix} 16 & -21.6 \\ -21.6 & 29.16 \end{bmatrix} \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}$$

$$S_\omega^{-1} S_B = \begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix}$$

Step - 4:

The LDA projection is then obtained as the solution of the generalized eigen value problem.

$$S_\omega^{-1} S_B V = \lambda V$$

$$| S_\omega^{-1} S_B - \lambda I | = 0$$

$$\begin{vmatrix} (11.89 - \lambda) & 8.81 \\ 5.08 & (3.76 - \lambda) \end{vmatrix} = 0$$

$$(11.89 - \lambda)(3.76 - \lambda) - (8.81)(5.08) = 0$$

$$\lambda = 15.85, 0$$

Step - 5:

$$w^* = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = S_\omega^{-1} [\mu_1 - \mu_2]$$

$$= \begin{bmatrix} 16 & -21.6 \\ -21.6 & 29.16 \end{bmatrix} \begin{bmatrix} -5.4 \\ -4 \end{bmatrix} = \begin{bmatrix} -0.91 \\ -0.39 \end{bmatrix}$$

$$\therefore \quad w_1 = -0.91$$
$$w_2 = -0.39$$