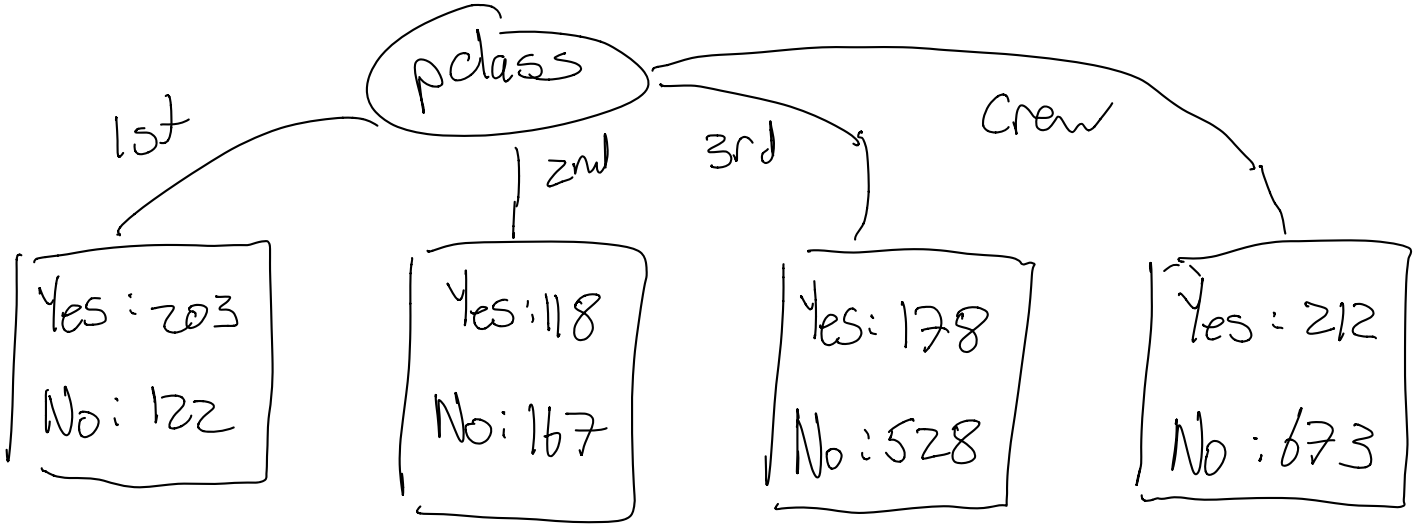
1.

Entropy Calculations

a)



pclass = 1st

$$info = ([203, 122]) = entropy\left(\frac{203}{325}, \frac{122}{325}\right)$$

$$= -\frac{203}{325}\log_2\left(\frac{203}{325}\right) - \frac{122}{325}\log_2\left(\frac{122}{325}\right) = 0.955$$

pclass = 2nd

$$info = ([118, 167]) = entropy\left(\frac{118}{285}, \frac{167}{285}\right)$$

$$= -\frac{118}{285}\log_2\left(\frac{118}{285}\right) - \frac{167}{285}\log_2\left(\frac{167}{285}\right) = 0.979$$

pclass = 3rd

$$\text{Info} = ([178, 528]) = \text{entropy}\left(\frac{178}{706}, \frac{528}{706}\right)$$

$$= -\frac{178}{706}\log_2\left(\frac{178}{706}\right) - \frac{528}{706}\log_2\left(\frac{528}{706}\right) = 0.815$$

pclass = Crew

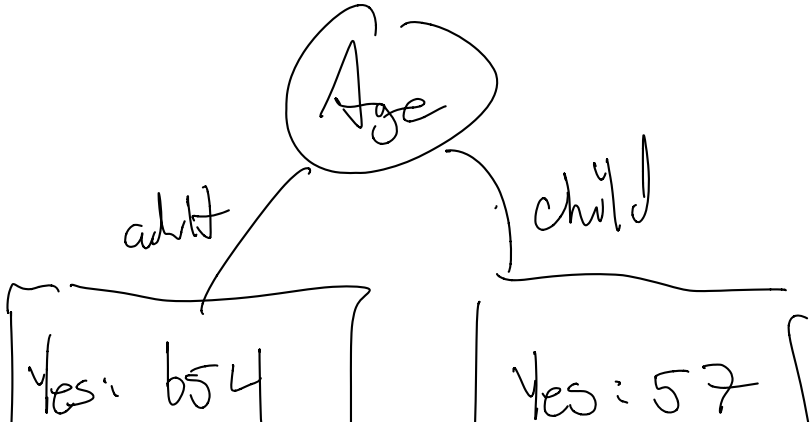
$$\text{Info} = ([212, 673]) = \text{entropy}(212, 673)$$

$$= -\frac{212}{885}\log_2\left(\frac{212}{885}\right) - \frac{673}{885}\left(\frac{673}{885}\right) = 0.794$$

Expected Info: pclass $= \text{Info}([203, 122], [118, 167], [178, 528], [212, 673])$

$$= 0.955\left(\frac{325}{2201}\right) + 0.979\left(\frac{285}{2201}\right) + 0.815\left(\frac{706}{2201}\right) + .794\left(\frac{885}{2201}\right)$$

$= .848$



No: 1438

No: 52

Age: adult

$$info([654, 1438]) = entropy(654, 1438)$$

$$= -\frac{654}{2092}\log_2\left(\frac{654}{2092}\right) - \frac{1438}{2092}\log_2\left(\frac{1438}{2092}\right) = 0.896$$

Age: child

$$info([57, 52]) = entropy(57, 52)$$

$$= -\frac{52}{109}\log_2\left(\frac{57}{109}\right) - \frac{52}{109}\log_2\left(\frac{52}{109}\right) = 0.998$$

Expected Info: Age $= info([654, 1438], [57, 52])$

$$= .896\left(\frac{2092}{2201}\right) + .998\left(\frac{109}{2201}\right)$$

$$= 0.901$$

Yes: 344

Yes: 367

No : 126

No : 1364

Sex : female

$$\text{info}([344,126]) = \text{entropy}(344,126)$$

$$= -\frac{344}{470}\log_2\left(\frac{344}{470}\right) - \frac{126}{470}\log_2\left(\frac{126}{470}\right) = 0,839$$

Sex : male

$$\text{info}([367,1364]) = \text{entropy}(367,1364)$$

$$= -\frac{367}{1731}\log_2\left(\frac{367}{1731}\right) - \frac{1364}{1731}\log_2\left(\frac{1364}{1731}\right) = 0,745$$

Expected Info : Sex $\quad \text{info}([344,126],[367,1364])$

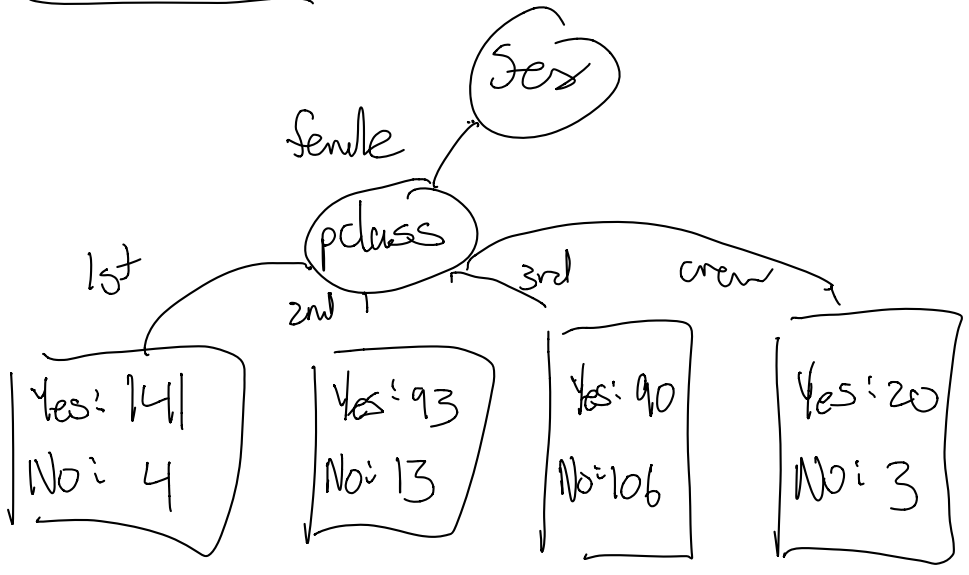
$$= .839\left(\frac{470}{2201}\right) + .745\left(\frac{1731}{2201}\right) \quad = 0765$$

- The lowest entropy for the 3 attributes is Sex

Root :

- testing 1st level possibilities
  - ↳ female → pclass
  - ↳ female → age
  - ↳ male → pclass
  - ↳ female → age

Female → pclass



1st : info([141,4]) = entropy(141,4)

$$= -\frac{141}{145}\log_2\left(\frac{141}{145}\right) - \frac{4}{145}\log_2\left(\frac{4}{145}\right) = 0,182$$

2nd: info([93,13]) = entropy(93,13)

$$= -\frac{93}{106}\log_2\left(\frac{93}{106}\right) - \frac{13}{106}\log_2\left(\frac{13}{106}\right) = 0,537$$

3rd: info([90,106]) = entropy(90,106)

$$= \frac{-90}{196} \cdot \log_2\left(\frac{90}{196}\right) - \frac{106}{196}\left(\frac{106}{196}\right) = 0,995$$

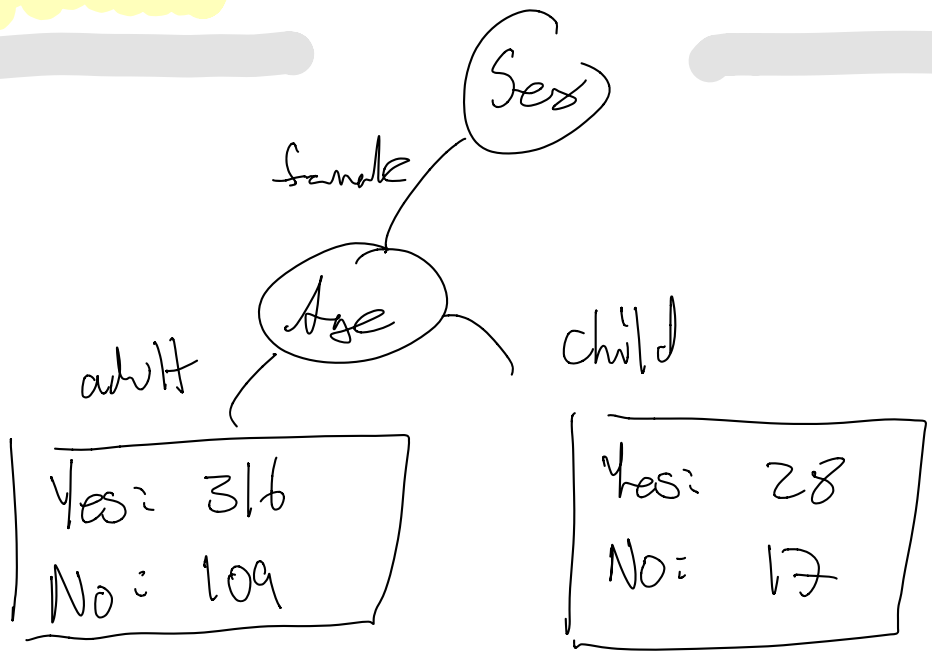
Crew: info([20, 3]) = entropy(20, 3)

$$= \frac{-20}{23} \log_2\left(\frac{20}{23}\right) - \frac{3}{23} \log_2\left(\frac{3}{23}\right) = 0,559$$

Expected Info =

$$= ,182\left(\frac{145}{470}\right) + ,537\left(\frac{106}{470}\right) + ,995\left(\frac{196}{470}\right) + ,559\left(\frac{23}{470}\right)$$

$= 0,620$



Age - Adult

info([316, 109]) = entropy(316, 109)

$$= -\frac{316}{425}\log_2\left(\frac{316}{425}\right) - \frac{104}{425}\log_2\left(\frac{104}{425}\right)$$

$$= 0,821$$

Age : Child

$$\text{info}([28,17]) = \text{entropy}(28,17)$$

$$= -\frac{28}{45}\log\left(\frac{28}{45}\right) - \frac{17}{45}\log_2\left(\frac{17}{45}\right)$$

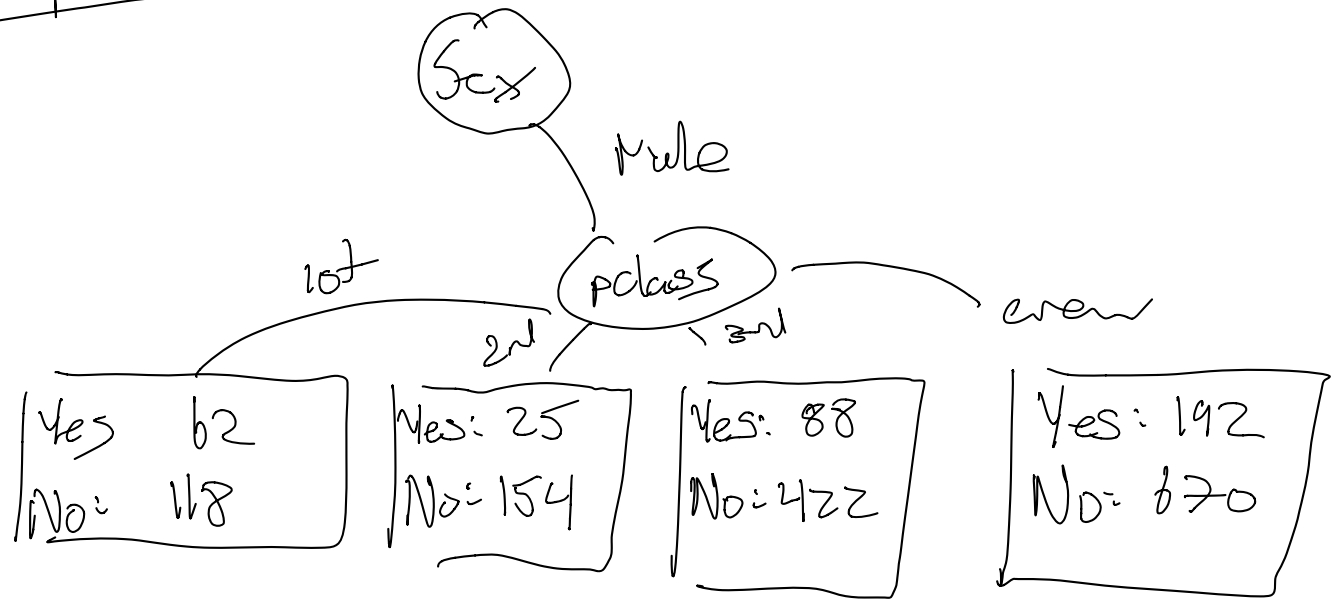
$$= 0,956$$

Expected info

$$= 0,821\left(\frac{425}{470}\right) + 0,956\left(\frac{45}{470}\right)$$

$$= 0,834$$

Rule → pclass



pclass: 1st

$$-\frac{62}{180}\log_2\left(\frac{62}{180}\right) - \frac{118}{180}\log_2\left(\frac{118}{180}\right) = 0.929$$

pclass: 2nd

$$-\frac{25}{179}\log_2\left(\frac{25}{179}\right) - \frac{154}{179}\log_2\left(\frac{154}{179}\right) = 0.583$$

pclass: 3rd

$$-\frac{88}{510}\log_2\left(\frac{88}{510}\right) - \frac{422}{510}\log_2\left(\frac{422}{510}\right) = 0.664$$

pclass: Crew

$$-\frac{192}{862}\log_2\left(\frac{192}{862}\right) - \frac{670}{862}\log_2\left(\frac{670}{862}\right) = 0.765$$

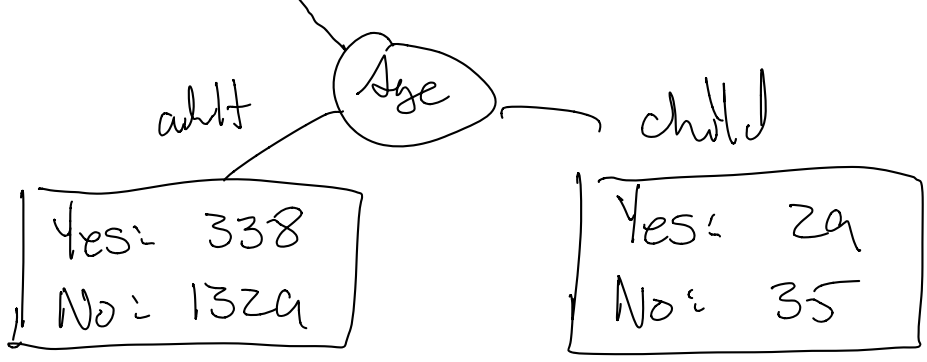
Expected:

$$.929\left(\frac{180}{1731}\right) + .583\left(\frac{179}{1731}\right) + .664\left(\frac{510}{1731}\right) + .765\left(\frac{862}{1731}\right)$$

$= 0.733$

Sex

male



Age: Adult

$$-\frac{338}{1667}\log_2\left(\frac{338}{1667}\right) - \frac{1329}{1667}\log_2\left(\frac{1329}{1667}\right) = 0.727$$

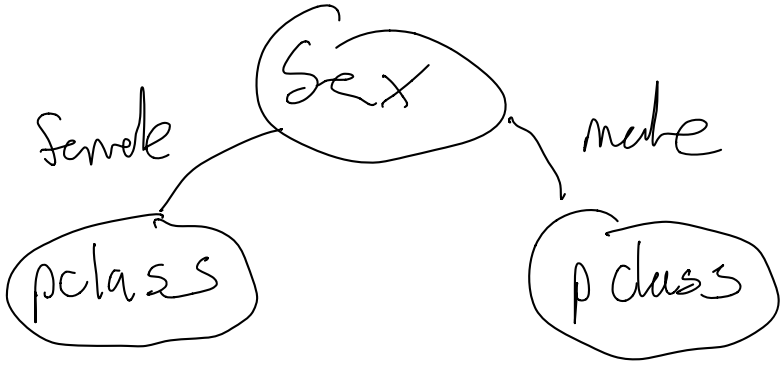
Age: Child

$$-\frac{29}{64}\log_2\left(\frac{29}{64}\right) - \frac{35}{64}\log_2\left(\frac{35}{64}\right) = 0.994$$

Expected Info:

$$.727\left(\frac{1667}{1731}\right) + .994\left(\frac{64}{1731}\right) = 0.737$$

- The lowest entropy is female → pclass and male → pclass so the tree becomes



- The results are affirmed in the attached q1.txt file from weka.

2. Following the PRISM Pseudo-Code

For each class C (ordered in ascendly order)
Initialize E to the instance set E is a set of instances of class C

While E contains instances of class C
- Create a rule R with an empty left-hand side that predicts class C
- Until R is perfect (or there are no more attributes to use) do
  - For each attribute A not mentioned in R, and each value v,
    - Consider adding the condition A=v to the left-hand side of R
    - Select A and v to maximize the accuracy p/t (break ties by choosing the condition with the largest t)
  - Add A=v to R
- Remove the instances covered by R from E

- Ascending order of class play = {no, yes}

If ?,

then recommendation = no

- possible tests

outlook = rainy	2/5
outlook = sunny	3/5 → we chose sunny
temp = cool	1/4
temp = mild	2/6
temp = hot	2/4
humidity = normal	1/7
humidity = high	4/7
windy = TRUE	3/6
windy = FALSE	2/8

- If outlook = sunny

and ?,

then recommendation = no

- possible tests

temp = hot	2/2
temp = mild	1/2
temp = cool	0/1
humidity = high	3/3 → we chose this

humidity = normal 0/2
windy = false 3/3
windy = true 1/2

1st Rule = If outlook = sunny
and humidity = high
then recommendation = no

(2)

If ?
then recommendation = no

outlook = rainy 2/5 → we choose this
temp = cool 1/4
temp = mild 2/5
humidity = normal 1/7
humidity = high 1/4
windy = true 2/5

if outlook = rainy
and ?
then recommendation = no.

tests

temp = cool 1/1
temp = mild 1/1

humidity = normal 1/1
humidity = high 1/1
windy = true 2/2 => choose this

2nd rule: if outlook = rainy
and windy = TRUE
then recommendation = yes

(3) if ?
then recommendation = yes

tests:

outlook = overcast 4/4 => choose
outlook = rainy 3/5
outlook = sunny 2/5
temp = hot 2/4
temp = cool 4/5
temp = mild 4/6
humidity = high 3/7
humidity = normal 6/7
windy = false 6/8
windy = true 3/6

3rd rule: if outlook = overcast
then recommendation = yes

(4) if :

then recommendation = yes

tests: outlook = rainy 3/5
outlook = sunny 2/5
temp = mild 3/5
temp = cool 2/3
humidity = high 1/5
humidity = normal 4/5 ⇒ choose
windy = false 4/6
windy = true 1/4

if humidity = normal
and :
then recommendation = yes

tests: outlook = rainy 2/3
outlook = sunny 2/2
temp = cool 2/3
temp = mild 2/2
windy = false 3/3 → choose
windy = true 1/2

4th rule: if humidity = normal
and windy = false

then recommendation = yes

(5) If :

then recommendation = yes

tests: outlook = rainy 1/3
outlook = sunny 1/4
temp = mild 2/4 ⇒ choose.
humidity = high 1/5

If temp = mild
and :
then recommendation = yes

tests: outlook = rainy 1/2
outlook = sunny 1/2
humidity = high 1/3
humidity = normal 1/1 ⇒ choose
windy = false 1/2
windy = true 1/2

5th Test:

If temp = mild
and humidity = normal
then recommendation = yes

(b) If ?

Then recommendation = yes

tests: outlook = rainy 1/3 ⇒ choose first one.
temp = mild 1/3
humidity = high 1/5
windy = false 1/3

If outlook = rainy
and ?
Then recommendation = yes

tests: temp = mild 1/3 ⇒ choose
humidity = high 1/5
windy = false 1/3

6th Test: if outlook = rainy
and temp = mild
Then recommendation = yes

- The full rule set

1st Rule = If outlook = sunny
and humidity = high

then recommendation = NO

2nd rule: if outlook = rainy
and windy = TRUE
then recommendation = yes

3rd rule: if outlook = overcast
then recommendation = yes

4th rule: if humidity = normal
and windy = false
then recommendation = yes

5th Test: if temp = mild
and humidity = normal
then recommendation = yes

6th Test: if outlook = rainy
and temp = mild
then recommendation = yes

↳ The q2.txt file provides similar results generated by weka.

3.)

pclass	age	sex	survived
2nd	child	male	?

survived
no = 1490, yes = 711
total = 2201

$P(\text{Survived} = \text{yes} \mid E) =$

$P(pclass = 2nd \mid survived = yes) *$
$P(age = child \mid survived = yes) *$
$P(sex = male \mid survived = yes) *$
$P(survived = yes) \;/\; P(E)$

$$= \frac{118/711 * 57/711 * 367/711 * 711/2201}{P(E)}$$

$$= \frac{0,00221852}{P(E)}$$

$P(Survived = no \mid E) =$

$P(pClass = 2nd \mid survived = no) *$
$P(age = child \mid survived = no) *$
$P(sex = male \mid survived = no) *$
$P(survived = no) \;/\; P(E)$

$$= \frac{87/1490 * 52/1490 * 1364/1490 * 1364/2201}{P(E)}$$

$$= \frac{0,00221903}{P(E)}$$

- $P(survived = yes \mid E) + P(survived = no \mid E) = 1$

$$\frac{0.00221852}{P(E)} + \frac{0.00221903}{P(E)} = 1$$

$$P(E) = 0.00221852 + 0.00221903$$
$$= 0.00443755$$

So, $P(\text{survived} = \text{yes} \mid E) = \frac{0.00221852}{0.00443755}$

$$= 50\%$$

$$P(\text{survived} = \text{no} \mid E) = \frac{0.00221903}{0.00443755}$$

$$= 50\%.$$

b)

pClass	age	sex	survived
2nd	adult	female	?