CSC 474 Assignment 1

Nik Rados V00801209 & Max Kasprzik V00722495

QUESTION 1

Class

Class	Yes	No	Total
1st	203	122	325
2nd	118	167	285
3rd	178	528	706
Crew	212	673	885

1st

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

$$= -$$

$$frac203325 \cdot log(\frac{203}{325}) - \frac{122}{325} \cdot log(\frac{122}{325}) \approx 0.95472$$

2nd

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

$$= -\frac{118}{285} \cdot log(\frac{118}{285}) - \frac{167}{285} \cdot log(\frac{167}{285}) \approx 0.97858$$

3rd

$$info([178, 528]) = entropy(\frac{178}{706}, \frac{528}{706}) =$$

$$= -\frac{178}{706} \cdot log(\frac{178}{706}) - \frac{528}{706} \cdot log(\frac{528}{706}) \approx 0.81463$$

Crew

$$info([212, 673]) = entropy(\frac{212}{885}, \frac{673}{885}) =$$

$$= -\frac{212}{885} \cdot log(\frac{212}{885}) - \frac{673}{885} \cdot log(\frac{673}{885}) \approx 0.79429$$

Average Weighted Entropy (Class)

$$info([203, 122], [118, 167], [178, 528], [212, 673]) =$$

$$= 0.95472 \cdot (\frac{325}{2201}) + 0.97858 \cdot (\frac{285}{2201}) + 0.81463 \cdot (\frac{706}{2201}) + 0.79429 \cdot (\frac{885}{2201}) \approx$$

$$\approx 0.84837$$

Age

Age	Yes	No	Total
Adult	654	1438	2092
Child	57	52	109

Adult

$$info([654, 1438]) = entropy(\frac{654}{2092}, \frac{1438}{2092}) =$$
$$= -\frac{654}{2092} \cdot log(\frac{654}{2092}) - \frac{1438}{2092} \cdot log(\frac{1438}{2092}) \approx 0.89617$$

Child

$$info([57, 52]) = entropy(\frac{57}{109}, \frac{52}{109}) =$$
$$= -\frac{57}{109} \cdot log(\frac{57}{109}) - \frac{52}{109} \cdot log(\frac{52}{109}) \approx 0.79429$$

Average Weighted Entropy (Age)

$$info([654, 1438], [57, 52]) =$$
$$= 0.89617 \cdot (\frac{654}{2201}) + 0.79429 \cdot (\frac{1438}{2201}) \approx$$
$$\approx 0.90124$$

Sex

Sex	Yes	No	Total
Male	367	1364	1731
Female	344	126	470

Male

$$info([367, 1364]) = entropy(\frac{367}{1731}, \frac{1364}{1731}) =$$
$$= -\frac{367}{1731} \cdot log(\frac{367}{1731}) - \frac{1364}{1731} \cdot log(\frac{1364}{1731}) \approx 0.74532$$

Female

$$info([344, 126]) = entropy(\frac{344}{470}, \frac{126}{470}) =$$
$$= -\frac{344}{470} \cdot log(\frac{344}{470}) - \frac{126}{470} \cdot log(\frac{126}{470}) \approx 0.83870$$

Average Weighted Entropy (Sex)

$$info([367, 1364], [344, 126]) =$$
$$= 0.74532 \cdot (\frac{1731}{2201}) + 0.83870 \cdot (\frac{470}{2201}) \approx$$
$$\approx 0.76526$$

Lowest Entropy = Sex = Root

Sex --> Male --> Class

Male - 1st Class

$$info([62, 118]) = entropy(\frac{62}{180}, \frac{118}{180}) =$$
$$= -\frac{62}{180} \cdot log(\frac{62}{180}) - \frac{118}{180} \cdot log(\frac{118}{180}) \approx 0.92901$$

Male - 2nd Class

$$info([25, 154]) = entropy(\frac{25}{179}, \frac{154}{179}) =$$
$$= -\frac{25}{179} \cdot log(\frac{25}{179}) - \frac{154}{179} \cdot log(\frac{154}{179}) \approx 0.76513$$

Male - 3rd Class

$$info([88, 422]) = entropy(\frac{88}{510}, \frac{422}{510}) =$$
$$= -\frac{88}{510} \cdot log(\frac{88}{510}) - \frac{422}{510} \cdot log(\frac{422}{510}) \approx 0.66350$$

Male - Crew

$$info([192, 670]) = entropy(\frac{192}{862}, \frac{670}{862}) =$$
$$= -\frac{192}{862} \cdot log(\frac{192}{862}) - \frac{670}{862} \cdot log(\frac{670}{862}) \approx 0.76513$$

Average Weighted Entropy (Male --> Class)

$$info([62, 118], [25, 154], [88, 422], [192, 670]) =$$
$$= 0.92901 \cdot (\frac{180}{1731}) + 0.58336 \cdot (\frac{179}{1731}) + 0.66350 \cdot (\frac{510}{1731}) + 0.76513 \cdot (\frac{862}{1731}) \approx$$
$$\approx 0.73349$$

Sex --> Male --> Age

Male - Adult

$$info([338, 1329]) = entropy(\frac{338}{1667}, \frac{1329}{1667}) =$$
$$= -\frac{338}{1667} \cdot log(\frac{338}{1667}) - \frac{1329}{1667} \cdot log(\frac{1329}{1667}) \approx 0.72741$$

Male - Child

$$info([29, 35]) = entropy(\frac{29}{64}, \frac{35}{64}) =$$
$$= -\frac{29}{64} \cdot log(\frac{29}{64}) - \frac{35}{64} \cdot log(\frac{35}{64}) \approx 0.99365$$

Average Weighted Entropy (Male --> Age)

$$info([338, 1329], [29, 35]) =$$
$$= 0.72741 \cdot (\frac{1667}{1731}) + 0.99365 \cdot (\frac{64}{1731}) \approx$$
$$\approx 0.73725$$

Lowest Entropy = Male --> Class

Sex --> Female --> Class

Female - 1st Class

$$info([141, 4]) = entropy(\frac{141}{145}, \frac{4}{145}) =$$
$$= -\frac{141}{145} \cdot log(\frac{141}{145}) - \frac{4}{145} \cdot log(\frac{4}{145}) \approx 0.18214$$

Female - 2nd Class

$$info([93, 13]) = entropy(\frac{93}{106}, \frac{13}{106}) =$$
$$= -\frac{93}{106} \cdot log(\frac{93}{106}) - \frac{13}{106} \cdot log(\frac{13}{106}) \approx 0.53691$$

Female - 3rd Class

$$info([90, 106]) = entropy(\frac{90}{196}, \frac{106}{196}) =$$
$$= -\frac{90}{196} \cdot log(\frac{90}{196}) - \frac{106}{196} \cdot log(\frac{106}{196}) \approx 0.99519$$

Female - Crew

$$info([20, 3]) = entropy(\frac{20}{23}, \frac{3}{23}) =$$
$$= -\frac{20}{23} \cdot log(\frac{20}{23}) - \frac{3}{23} \cdot log(\frac{3}{23}) \approx 0.55863$$

Average Weighted Entropy (Female --> Class)

$$info([141, 4], [93, 13], [90, 106], [20, 3]) =$$
$$= 0.18214 \cdot (\frac{145}{470}) + 0.53691 \cdot (\frac{106}{470}) + 0.99519 \cdot (\frac{196}{470}) + 0.55863 \cdot (\frac{23}{470}) \approx$$
$$\approx 0.61964$$

Sex --> Female --> Age —> test

Female - Adult

$$info([316, 109]) = entropy(\frac{316}{425}, \frac{109}{425}) =$$
$$= -\frac{316}{425} \cdot log(\frac{316}{425}) - \frac{109}{425} \cdot log(\frac{109}{425}) \approx 0.82137$$

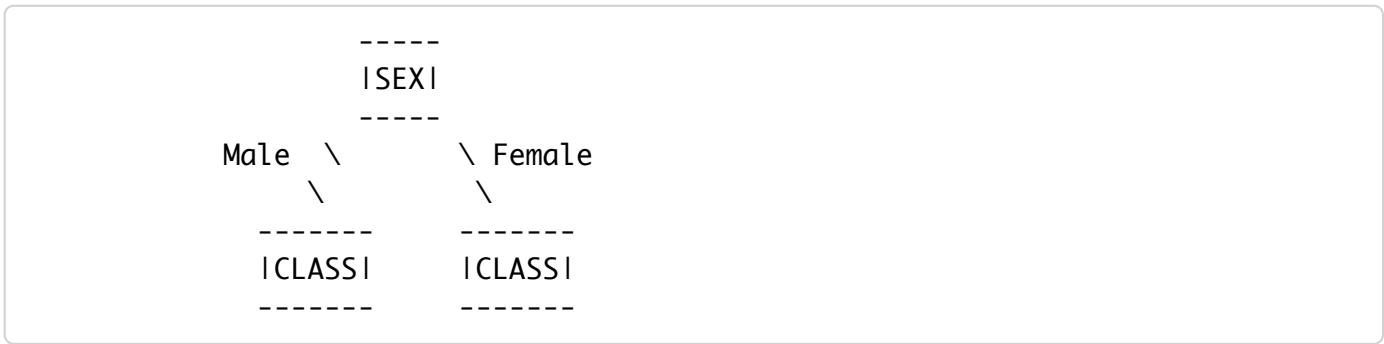
Female - Child

$$info([28, 17]) = entropy(\frac{28}{45}, \frac{17}{45}) =$$
$$= -\frac{28}{45} \cdot log(\frac{28}{45}) - \frac{17}{45} \cdot log(\frac{17}{45}) \approx 0.95646$$

Average Weighted Entropy (Female --> Age)

$$info([316, 109], [28, 17]) =$$
$$= 0.82137 \cdot (\frac{425}{470}) + 0.95646 \cdot (\frac{45}{470}) \approx$$
$$\approx 0.83430$$

Lowest Entropy = Female --> Class



QUESTION 2

We are looking for **If ? then play = yes** where:

$$\frac{P}{t} = 1$$

Rule 1

Dataset

Classes	Attribute	Play = Yes / Total
Outlook	**Overcast**	**4/4**
	Sunny	2/5
	Rainy	3/5
Temperature	Hot	2/4
	Mild	4/6

Classes	Attribute	Play = Yes / Total
	Cool	3/4
Humidity	High	3/7
	Normal	6/7
Windy	True	3/6
	False	6/8

Rule 1 => If outlook = overcast then play = yes

Rule 2

Dataset

Classes	Attribute	Play = Yes / Total
Outlook	Sunny	2/5
	Rainy	3/5
Temperature	Hot	0/2
	Mild	3/5
	Cool	2/3
Humidity	High	1/5
	Normal	**4/5**
Windy	True	1/4
	False	4/6

So far we have, **If humidity = normal and ? then play = yes**

Narrow Dataset to Humidity = Normal

Classes	Attribute	Play = Yes / Total
Outlook	Sunny	2/2
	Rainy	2/3
Temperature	Hot	1/1
	Mild	2/2
	Cool	3/4
Windy	True	2/3

Classes	Attribute	Play = Yes / Total
	False	4/4

Rule 2 = If humidity = normal and windy = false then play = yes

QUESTION 3

Evidence = E = 2nd, child, male

$$P(survive = yes|E) = \frac{(P(2nd|yes) \cdot P(child|yes) \cdot P(male|yes) \cdot P(yes))}{P(E)}$$
$$= \frac{(\frac{118}{711} \cdot \frac{57}{711} \cdot \frac{367}{711} \cdot \frac{711}{2201})}{P(E)}$$
$$= \frac{0.0022}{P(E)}$$

$$P(survive = no|E) = \frac{(P(2nd|no) \cdot P(child|no) \cdot P(male|no) \cdot P(no))}{P(E)}$$
$$= \frac{(\frac{167}{1490} \cdot \frac{52}{1490} \cdot \frac{1364}{1490} \cdot \frac{1490}{2201})}{P(E)}$$
$$= \frac{0.0024}{P(E)}$$

Normalization Constant

$$P(survive = yes|E) + P(survive = no|E) = 1$$
$$= \frac{0.0022}{P(E)} + \frac{0.0024}{P(E)} = 1$$
$$P(E) = (0.0022 + 0.0024)$$

Therfore,

$$P(survive = yes|E) = \frac{0.0022}{(0.0022 + 0.0024)} = 0.478 = 47.8\%$$

$$P(survive = no|E) = \frac{0.0024}{(0.0022 + 0.0024)} = 0.522 = 52.2\%$$