

SENG 474
Assignment 1
Avery Kushner V00865118
Derek Siemens V008720720

Question 1

The following data was calculated by entropy.py
Consult QuestionOneWekaResults.txt for weka result

ROOT DATA

*** class entropies***

{'1st': 0.9547169960042388, '2nd': 0.9785707523227567, '3rd': 0.814628274498826, 'crew':
0.7942878485408212}
total class entropy 0.8479781997584745

*** age entropies***

{'Adult': 0.8961741195225359, 'Child': 0.998481607797779}
total age entropy 0.9008314047643521

*** sex entropies***

{'Male': 0.7453189521844144, 'Female': 0.8387034444830612}
total sex entropy 0.764912681715831
Lowest Average Entropy, continue with sex as root

MALE DATA

*** male-age entropies***

{'adult': 0.7274127790215612, 'child': 0.9936507116910405}
total male-age entropy 0.7372563536552104

*** male-class entropies***

{'1st': 0.9290083681703217, '2nd': 0.5833602911671452, '3rd': 0.6635027700460849, 'crew':
0.7651352640546407}
total male-class entropy 0.7334350137077876
Lowest Average Entropy

FEMALE DATA

*** female-age entropies***

{'adult': 0.8213736008494182, 'child': 0.9564574047992596}

total female-age entropy 0.8343071565467435

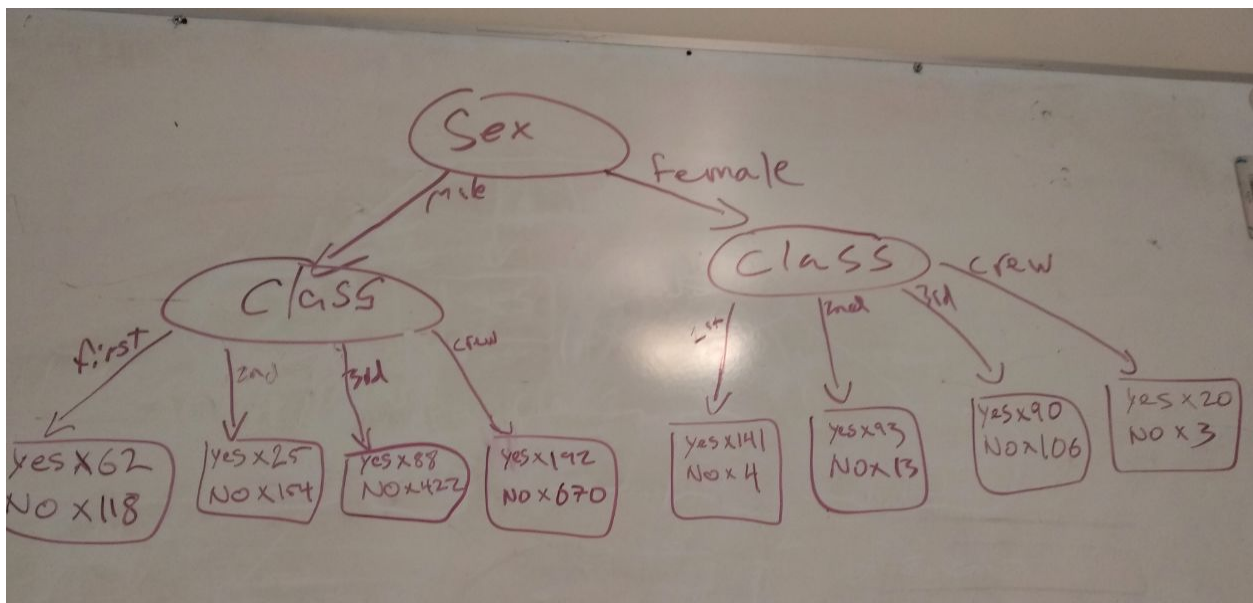
*** female-class entropies***

{'1st': 0.18213846457886657, '2nd': 0.5369064378756414, '3rd': 0.9951876662918927, 'crew': 0.5586293734521992}

total female-class entropy 0.6196328041731174

Lowest Average Entropy

Final Decision Tree



Question 2

Consult QuestionTwoWekaResult.txt for weka result

outlook	temperature	humidity	windy	play				
sunny	hot	high	FALSE	no				
sunny	hot	high	TRUE	no				
overcast	hot	high	FALSE	yes				
rainy	mild	high	FALSE	yes				
rainy	cool	normal	FALSE	yes				

rainy	cool	normal	TRUE	no				
overcast	cool	normal	TRUE	yes				
sunny	mild	high	FALSE	no				
sunny	cool	normal	FALSE	yes				
rainy	mild	normal	FALSE	yes				
sunny	mild	normal	TRUE	yes				
overcast	mild	high	TRUE	yes				
overcast	hot	normal	FALSE	yes				
rainy	mild	high	TRUE	no				
Constructing First Rule								
If ? Then Play=Yes					If Outlook= Overcast and ? Then Play=Yes			
			Clearly Outlook= Overcast is the winner	continued...				Covers all occurrences
Outlook=...	Sunny	2/5			Temperature=...	Hot	2/2	
	Overcast	4/4				Mild	1/1	
	Rainy	3/5				Cool	1/1	
Temperature=.. .	Hot	2/4			Humidity=...	High	2/2	
	Mild	4/6				Normal	2/2	
	Cool	3/4						
					Windy=...	TRUE	2/2	
Humidity=...	High	3/7				FALSE	2/2	
	Normal	6/7						
Windy=...	TRUE	3/6		FIRST RULE:	If Outlook= Overcast Then Play=Yes			
	FALSE	6/8						

Constructing Second Rule								
^^^^REFER TO THE STATS ABOVE ^^^^								
The second best percentage is Humidity= Normal								
Thus								
If Humidity= Normal and ? Then Play=Yes								
				Although there are many options here with 100%, Windy= FALSE is the 100% option with the highest coverage				
Outlook=...	Sunny	2/2		Thus our second rule is...				
	Rainy	2/3						
	Overcast	2/2		SECOND RULE:	If Humidity= Normal and Windy= FALSE then Play=Yes			
Temperature=.. .	Hot	1/1						
	Mild	2/2						
	Cool	3/4						
Windy=	TRUE	2/3						

	FALSE	4/4						
To Recap, our two rules are as follows:								
FIRST RULE:	If Outlook= Overcast Then Play=Yes							
SECOND RULE:	If Humidity= Normal and Windy= FALSE then Play=Yes							

Question 3

Wrote and used naive_bayes.py script to compute the following output:

2nd child male ?

$P(\text{survived} | E) \text{ alpha} * 0.0017398401599753273$

$P(\text{didn't survive} | E) \text{ alpha} * 0.0024240501791887303$

Our Alpha Value is $1/P(\text{survived} | E) + P(\text{didn't survive} | E) = 240.160023090512\%$

Probability of yes: 41.78400529934747%

Probability of no: 58.215994700652516%

Thus classified as No

2nd adult female ?

$P(\text{survived} | E) \text{ alpha} * 0.018711328449285785$

$P(\text{didn't survive} | E) \text{ alpha} * 0.0024240501791887303$

Our Alpha Value is $1/P(\text{survived} | E) + P(\text{didn't survive} | E) = 47.31403291033338\%$

Probability of yes: 88.53084100455649%

Probability of no: 11.469158995443511%

Thus classified as Yes

Question 4

Refer to results.txt which was computed using naive_bayes_text_classification.py