בוחן 2017 – פתרון החלק החישובי

<u>סעיף א'</u>

נתונים:

		Class (Buy car)	
Sunroof	**Air Bags**	**No**	**Yes**
0	**0**	**10**	**5**
0	**1**	**5**	**10**
1	**2**	**5**	**0**
1	**0**	**0**	**40**
0	**1**	**30**	**20**
0	**2**	**15**	**15**
1	**0**	**10**	**5**
1	**1**	**25**	**5**

תשובות סופיות:

	Sunroof	Air bags	Class
Entropy / Split Info	0.993	1.480	1.000
Cond. Entropy	0.993	0.925	
Info. Gain	0.007	0.075	
Info. Gain Ratio	0.007	0.050	

ספירת מקרים עבור sunroof:

sunroof	No	Yes	TOTAL
0	60	50	110
1	40	50	90
TOTAL	100	100	200

ספירת מקרים עבור air bag:

AIR BAG	No	Yes	TOTAL
0	20	50	70
1	60	35	95
2	20	15	35
TOTAL	100	100	200

חישוב השורה הראשונה בטבלה:

Split info (sunroof) = $-(110/200)*\log_2(110/200)-(90/200)*\log_2(90/200)=0.993$

Split info (air bag) = $-(70/200)*\log_2(70/200)-(95/200)*\log_2(95/200)-(35/200)*\log_2(35/200)=1.480$

Entropy (class) = $-(100/200)*\log_2(100/200)-(100/200)*\log_2(100/200)=1$

חישוב השורה השנייה בטבלה:

Cond entropy (sunroof) = (110/200)*[-(60/110)*$\log_2$(60/110)-(50/110)* $\log_2$(50/110)]+(90/200)*[-(40/90)*$\log_2$(40/90)-(50/90)*$\log_2$(50/90)]=0.993

Cond entropy (air bag) = (70/200)*[-(20/70)*$\log_2$(20/70)-(50/70)* $\log_2$(50/70)]+(95/200)*[-(60/95)*$\log_2$(60/95)-(35/95)*$\log_2$(35/95)]+(35/200)*[-(20/35)*$\log_2$(20/35)-(15/35)*$\log_2$(15/35)]=0.925

חישוב השורה השלישית בטבלה:

Information gain (sunroof) = entropy(class) – cond entropy (sunroof) = 1 – 0.993 = 0.007

Information gain (air bag) = entropy (class) – cond entropy (air bag) = 1 – 0.925 = 0.075

חישוב השורה הרביעית בטבלה:

Gain ratio (sunroof) = information gain (sunroof) / split info (sunroof) = 0.007/0.993 = 0.007

Gain ratio (air bag) = information gain (air bag) / split info (air bag) = 0.075/1.480 = 0.050

נבחר לפצל את קודקוד השורש לפי המשתנה air bag מכיוון שלו יש את ערך ה-gain ratio הגבוה יותר (0.050 עבור air bag לעומת 0.007 עבור sunroof).

עץ תקין צריך להכיל קודקוד שורש, פיצול לפי המשתנה air bag עם ציון ערכי הפיצול השונים (0, 1, 2), ועלי העץ צריכים להכיל את התפלגות משתנה המטרה.

<u>סעיף ב'</u>

יש לחשב pessimistic error עבור העץ שנבחר בסעיף א' לפני ואחרי גיזום העץ.

Pessimistic error before pruning:

q(T) = [(20+35+15) + (0.5)*3]/200 = 71.5/200 = 0.357

pessimistic error after pruning:

q(v) = (100+0.5)/200 = 100.5/200 = 0.502

במידה ונעשתה טעות ובסעיף א' נבחר המשתנה sunroof:

Pessimistic error before pruning:

q(T) = [(50+40) + (0.5)*2]/200 = 91/200 = 0.455