

חלק 2 [50 נקודות]

- יש להציג את כל התוצאות עם שלוש ספרות אחרי נקודה עשרונית אלא אם צוין אחרת!
- יש לרשום את כל התשובות על-גבי שאלון הבחינה בלבד
- טיוטות החישוב ייגרסו ללא בדיקה

קרא בעיון את השאלה ואת הנתונים הבאים.

נתון בסיס הנתונים הבא:

Purchase Vehicle Dataset

First Hand	Sunroof	Air bags	Class (Buy car)	
			No	Yes
No	No	0	10	5
No	No	1	5	10
No	Yes	2	5	0
No	Yes	0	0	40
Yes	No	1	30	20
Yes	No	2	15	15
Yes	Yes	0	10	5
Yes	Yes	1	25	5

Attribute	Type	Use
First Hand	Nominal	Input
Sunroof	Nominal	Input
Air bags	Numeric	Input
Class	Nominal	Target

(טבלת ערכי חי בריבוע ברמת-מובהקות של 95%):

DF	1	2	3	4	5	6
Chi Square (95%)	3.841	5.991	7.815	9.488	11.070	12.592

שימו לב:

- Class מייצג עבור כל אחד מהסיווגים האפשריים של: yes, no את מספר המופעים בבסיס הנתונים. לדוגמא, עבור השורה הראשונה בטבלה, יש 10 רשומות שמסווגות ל-no ו-5 רשומות שמסווגות ל-yes. יש לענות על כל הסעיפים הבאים:

א. יש לבנות את הרמה הראשונה ברשת אינפו-עמומה (IFN) עבור בסיס הנתונים המסומם "Purchase Vehicle" תוך שימוש ברמת בטחון של 95%. עבור כל משתנה יש להציג את חישובי הביניים ואת התוצאות הסופיות

הבאות: שם המאפיין, ערכי פיצול (אם רלבנטי), χ^2 , G^2 , MI ומסקנה (האם דוחים/לא דוחים את השערת האפס).

סמן בעיגול את המשתנה הנבחר לפיצול של השכבה בטבלת התוצאות הסופיות. 20 נקודות

חישובי ביניים

First Hand

	כמות תצפיות No	Cond. Prob.	Joint Prob.	כמות תצפיות Yes	Cond. Prob.	Joint Prob.	Total	Cond. Prob.
Total								

Sunroof

	כמות תצפיות No	Cond. Prob.	Joint Prob.	כמות תצפיות Yes	Cond. Prob.	Joint Prob.	Total	Cond. Prob.
Total								

Air bags

	כמות תצפיות No	Cond. Prob.	Joint Prob.	כמות תצפיות Yes	Cond. Prob.	Joint Prob.	Total	Cond. Prob.
Total								

Air bags

	כמות תצפיות No	Cond. Prob.	Joint Prob.	כמות תצפיות Yes	Cond. Prob.	Joint Prob.	Total	Cond. Prob.
Total								

Air bags

	כמות תצפיות No	Cond. Prob.	Joint Prob.	כמות תצפיות Yes	Cond. Prob.	Joint Prob.	Total	Cond. Prob.
Total								

התוצאות הסופיות

מסקנה (לדחות / לא לדחות H_0)	χ^2	G^2	MI	ערכי פיצול	מאפיין

ב. חשבו את המשקולות של הצלעות עבור המודל מסעיף א' והציגו בצורה גרפית את הרשת. 10 נקודות

No.	Rule	Weight

גרף:



ג. חשבו את שגיאת האימון של המודל מסעיף ב'. 5 נקודות

שגיאת אימון (%)	
-----------------	--



ד. נניח כי המאפיינים של בסיס הנתונים "Purchase Vehicle" מייצגים טרנזקציות, כלומר, אם הערך של המאפיין הוא "Yes" או גדול מ-0, סימן שהמאפיין מופיע בטרנזקציה. יש להתחשב רק בטרנזקציות שמכילות פריט אחד לפחות!

יש להשתמש באלגוריתם ה-Apriori ולהחזיר את כל החוקים של שני פריטים ומעלה אשר מקיימים את

האילוצים: $\text{min support} = 20\%$, $\text{min confidence} = 40\%$

15 נקודות

Rule	Support	Confidence