

חלק ב' [50 נקודות]

יש להציג את כל התוצאות עם שלוש ספרות אחרי נקודה עשרונית אלא אם צוין אחרת!

נתונה טבלת נתונים של לקוחות בחברה סלולארית. מטרת הטבלה לחזות את הסתברות הנטישה (churn) של לקוח.

מס' רשומה	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
דקות שיחה בשנה האחרונה	3216	3156	2952	0	2912	164	2060	2892	2044	0	0	1204	292	228	3560
נטישה	לא	לא	לא	לא	לא	לא	לא	לא	לא	לא	כן	כן	כן	כן	כן

התפלגות חי בריבוע (מובהקות של 90%)

DF	1	2	3	4	5	6
Chi Square (90%)	2.706	4.605	6.251	7.779	9.236	10.645

יש לבנות רשת אינפּו-עמומה (IFN) עבור טבלת הנתונים הנ"ל תוך שימוש ברמת ביטחון של 90%. עליכם לבחון אִךְ ורק את נקודות הפיצול הבאות של המשתנה "דקות שיחה בשנה האחרונה": 228, 1204, 2044, 3156.

א. יש לפרט את חישובי האלגוריתם בטבלה הבאה 18 נקודות:

כמות תצפיות לפני הפיצול	האינטרוואל הראשון (מ- עד-)	האינטרוואל השני (מ- עד-)	הסתברות הנטישה באינטרוואל הראשון	הסתברות הנטישה באינטרוואל השני	האינפורמציה ההדדית המותנית	$G^2$	לפצל (כן / לא)



ב. יש להציג את הרשת שבניתם בצורה גראפית תוך מספור כל הקודקודים וציון הסיווג החזוי בכל קודקוד טרמינלי (מבלי לחרוג מגבולות המסגרת!) **10 נקודות** :

[illegible]

ג. יש לחשב את משקולות החיבור (connection weights) של כל החוקים המתקבלים מהרשת 8 נקודות:

מס' חוק	מס' קודקוד טרמינלי	דקות שיחה בשנה האחרונה (מ- עד-)	סיווג נטישה (כן / לא)	הסתברות הסיווג	משקולת החיבור
					סה"כ משקולות

ד. יש לחשב את רווח בר-סמך לדיוק המודל IFN ברמת-ביטחון של 95% ( $z_{\alpha} = 1.96$ ) **4 נקודות.**

Lower Bound of Accuracy	Upper Bound of Accuracy

יש לחלק את כל הערכים של המשתנה "דקות שיחה בשנה האחרונה" לשלושה אינטרוואלים בעלי עומק שווה (equal depth) ולהחליק אותם לפי הממוצע של כל אינטרוואל. **בעמודה האחרונה** יש לחשב את סטיית התקן של הערכים המוחלקים. **10 נקודות.**

מס' רשומה	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	סטיית תקן
דקות שיחה בשנה האחרונה																