חלק ב' [60 נקודות]

נתונה טבלת סיכום של נתוני עובדים. העמודה "count" בטבלה מציינת את מספר העובדים בעלי מאפיינים מסוימים (מחלקה + סטאטוס + קבוצת גיל + משכורת). למשל, עפ"י השורה הראשונה בטבלה, 30 עובדים שייכים למחלקת המכירות (sales), נמצאים בסטאטוס בכיר (senior), הם בני 31-35 ומקבלים משכורת בין 46k ל-50k בשנה. משתנה המטרה הוא "סטאטוס העובד" (Status).

Department	Status	Age	Salary	Count
Sales	Senior	31..35	46k..50k	30
Sales	Junior	26..30	26k..30k	40
Sales	Junior	31..35	31k..35k	40
Systems	Junior	21..25	46k..50k	20
Systems	Senior	31..35	66k..70k	5
Systems	Junior	26..30	46k..50k	3
Systems	Senior	41..45	66k..70k	3
Marketing	Senior	36..40	46k..50k	10
Marketing	Junior	31..35	41k..45k	4
Secretary	Senior	46..50	36k..40k	4
Secretary	Junior	26..30	26k..30k	6

א. יש לחשב את האנטרופיה המותנית של סטאטוס העובד בהינתן אחד המשתנים הבאים: מחלקה, קבוצת גיל או משכורת. יש לסמן תשובה אחת בלבד **בכל שורה**. **15 נקודות**.

Department	0.049	0.497	0.850	0.899	⬅ תשובה
Age	0.047	0.474	0.486	0.540	⬅ תשובה
Salary	0.362	0.380	0.501	0.538	⬅ תשובה

ב. מהו דיוק "חוק הרוב" על נתוני האימון הנ"ל? יש לסמן תשובה אחת בלבד. **5 נקודות.**

0.500	0.685	0.805	1.000	⬅ תשובה

ג. מהם הסיווגים החזויים בעץ החלטה בעל רמה אחת בלבד המבוססת על אחד המשתנים הבאים: מחלקה, קבוצת גיל או משכורת? יש לסמן שורה אחת בלבד **בכל טבלה**. **15 נקודות.**

Department: תשובה ⬇	Sales	Systems	Marketing	Secretary
1	Senior	Senior	Senior	Junior
2	Senior	Senior	Junior	Junior
3	Senior	Junior	Junior	Senior
4	Junior	Junior	Senior	Junior

Age: תשובה ⬇	21..25	26..30	31..35	36..40	41..45	46..50
1	Junior	Junior	Senior	Senior	Senior	Senior
2	Junior	Junior	Junior	Senior	Senior	Senior
3	Senior	Senior	Junior	Junior	Junior	Junior
4	Senior	Junior	Junior	Senior	Senior	Senior

Salary: תשובה ⬇	26k..30k	31k..35k	36k..40k	41k..45k	46k..50k	66k..70k
1	Junior	Junior	Senior	Junior	Senior	Senior
2	Junior	Junior	Junior	Senior	Senior	Senior
3	Junior	Junior	Senior	Senior	Senior	Senior
4	Junior	Junior	Junior	Senior	Junior	Senior

ד. יש לחשב את דיוק האימון של עץ החלטה בעל רמה אחת בלבד המבוססת על אחד המשתנים הבאים: מחלקה, קבוצת גיל או משכורת. יש לסמן תשובה אחת בלבד **בכל שורה. 15 נקודות.**

Department	0.685	0.709	0.721	0.788	⬅ תשובה
Age	0.733	0.788	0.807	0.823	⬅ תשובה
Salary	0.758	0.861	0.904	0.953	⬅ תשובה

ה. יש לסווג את התצפיות הבאות באמצעות האלגוריתם k השכנים הקרובים ביותר (k-NN) תוך שימוש בהנחות הבאות (**10 נקודות**):

1) k=1

2) כל המשתנים נומינליים והמרחק ביניהם נמדד באמצעות "התאמה פשוטה" (Simple matching).

3) כל שורה בטבלת הנתונים הנ"ל מייצגת תצפית אחת בלבד (count = 1).

יש לסמן תשובה אחת בלבד **בכל שורה**

Record	Department	Age	Salary	Senior	Junior	שני הסיווגים	
1	Sales	31..35	26k..30k				⬅ תשובה
2	Sales	26..30	26k..30k				⬅ תשובה
3	Systems	31..35	66k..70k				⬅ תשובה
4	Marketing	36..40	46k..50k				⬅ תשובה
5	Marketing	31..35	66k..70k				⬅ תשובה