האוניברסיטה 🖧	הפתוחה מס־שא	520 - אלון
י"ב בתמוז תשע"ג	20	
	ביוני 2013	81
סמסטר 2013ב		מס' מועד
20595/4		
	שאלון בחינת גמר	מספר התלמיד הנבחן רשום את כל תשע הספרות
	20595 - כריית מידע	
	משך בחינה: 3 שעות	
מבנה הבחינה:	בשאלון זה <b>4</b> עמודים	
בבחינה שני חלקים.		
יש לענות על כל הש	ולות.	
חלק א: ענו על ארבע השאל 90 נקודות לחלק זר		
חלק ב: ענו על השאלה שבו 10 נקודות לחלק זר		
משקל כל שאלה מכ אפשר להשתמש בו פרטו את הנחותיכם	רט בגוף השאלון. נחות במהלך הפתרון, אם תזדקקו לכך.	
חומר עזר כל חומר עזר	מותר בשימוש.	

# השימוש במחשב כלשהו ו/או טלפון חכם אסור בשימוש. בהצלחה !!! החזירו למשגיח את השאלון

# חלק א (90 נקודות)

בחלק זה ארבע שאלות. ענו במפורט על ארבע השאלות.

שאלה 1 (30 נקודות) – עצי החלטה (decision trees)

נתונה טבלת נתוני אימון:

קידום	תואר אקדמי	משך זמן עבודה	מצב	מגדר	מסי
בעבודה		בארגון	משפחתי		סידורי
ל	ל	רב	נשוי	7	1
ל	כ	רב	נשוי	)	2
٥	ל	בינוני	נשוי	7	3
٥	ל	מועט	רווק	7	4
٥	ל	מועט	רווק	7	5
ל	כ	מועט	רווק	)	6
٥	כ	בינוני	רווק	)	7
ל	ל	רב	נשוי	7	8
٥	ל	רב	רווק	7	9
٥	ל	מועט	נשוי	7	10
٥	כ	רב	נשוי	)	11
٥	כ	בינוני	נשוי	)	12
٥	ל	בינוני	נשוי	7	13
ל	٥	מועט	נשוי	ì	14

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, עבור נתוני האימון שבטבלה לחיזוי החלטה האם העובד יקודם במקום עבודתו. בתשובתכם הדגימו את שלבי בחירת התכונה המפצלת בעץ.

הערה: יש לכלול חישוב של אחד המדדים כדוגמת אנטרופיה, Gain ratio, מדד גיני.

ב. איזה מבין התכונות ניתן להסיר ומדוע? אם אין תכונה הניתנת להסרה, יש לציין זאת מפורשות.

# המשך הבחינה בעמוד הבא

# (association rules) – חוקי הקשר (30 נקודות) – חוקי

: נתונה הטבלה הבאה

רישום לחוגים	מספר תלמיד
{כדורסל, כדורגל, שחמט}	1
{פסנתר, דרמה, כדורסל}	2
{כדורסל, שחמט, אומנות}	3
{פסנתר, דרמה }	4
{כדורגל, אומנות, פסנתר}	5
{כדורסל, שחמט, פסנתר, דרמה}	6
{פסנתר, אומנות, כדורסל}	7
{כדורגל, שחמט}	8
{כדורסל, שחמט, פסנתר, דרמה}	9
{כדורגל, אומנות}	10

# :בהנחה ש

Min\_support=50%

Min\_confidence=70%

מצאו את כל הקבוצות התדירות תוך שימוש באלגוריתם א-פריורי.

# (similarity and distance measures) שאלה 3 (סו נקודות) – מדדי מרחק ודימיון

: נתונים שני וקטורים בינאריים

X=0101010001

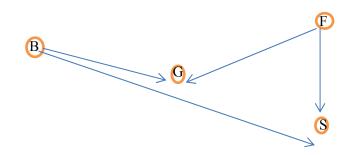
Y=0100011000

. Y-ל X בין (Jaccard similarity) בין (Hamming distance) את מרחק המינג (חשבו את מרחק ודמיון את המינג

# שאלה 4 (20 נקודות) – למידה בייאסינית

באיור הבא נתונה רשת בייסיאנית המתארת התנעת רכב:

התנעת הרכב (יסומן (F)) נקבעת בהתאם לכמות הדלק במיכל הדלק (יסומן (F)) ו/או למצב המצבר (יסומן (F)).



P(B=bad)=0.1

P(F = empty) = 0.2 : empty

: נגדיר

P(G=empty|B=good, F=not empty)=0.1

P(G=empty|B=good, F=empty)=0.8

P(G=empty|B=bad, F=not empty)=0.2

P(G=empty|B=bad, F=empty)=0.9

P(S=no|B=good, F=not empty)=0.1

P(S=no|B=good, F=empty)=0.8

P(S=no|B=bad, F=not empty)=0.9

P(S=no|B=bad, F=empty)=1.0

- א. בהנחה שהמצבר התרוקן (bad), חשבו את ההסתברות שהרכב יתניע.
  - ב. חשבו את ההסתברות (P(B=good,F=empty,G=empty,S=yes).

# חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

## שאלה 5 (10 נקודות)

במהלך קורס כריית מידע טען אחד הסטודנטים: ״בכריית חוקי הקשר משתמשים בשני מדדים – מדד תמיכה ומדד הביטחון. ואולם, מאחר וקיים קשר בין שני המדדים, הרי שניתן להשתמש רק במדד אחד בלבד.״

חוו דעתכם ונמקו את תשובתכם.

### בהצלחה!