

רשום את כל תשע הספרות

האוניברסיטה

כ"ו בתמוז תשע"ה

הדבק כאן את מדבקת הנבחן

מס' שאלון - 506

ביולי 2015

13

85 מס' מועד

סמסטר 2015ב

20595 / 4

שאלון בחינת גמר

20595 - כריית מידע

משך בחינה: שעות

> בשאלון זה 5 עמודים

מבנה הבחינה:

בבחינה שני חלקים.

יש לענות על כל השאלות.

:חלק א

ענו על ארבע השאלות שבחלק זה.

90 נקודות לחלק זה.

חלק ב:

ענו על השאלה שבחלק זה.

10 נקודות לחלק זה.

משקל כל שאלה מפורט בגוף השאלון.

אפשר להשתמש בהנחות במהלך הפתרון, אם תזדקקו לכך.

פרטו את הנחותיכם!

חומר עזר:

כל חומר עזר מותר בשימוש.

אסור בשימוש כל מכשיר שבאמצעותו אפשר להתחבר לאינטרנט

או לאצור מידע לרבות מחשב נישא ו/או טאבלט.

בהצלחה !!!

החזירו

למשגיח את השאלון

וכל עזר אחר שקיבלתם בתוך מחברת התשובות



# חלק א (90 נקודות)

בחלק זה ארבע שאלות. ענו במפורט על ארבע השאלות.

# (decision tree) עצי החלטה – (30 נקודות) – עצי

נתונה טבלת נתוני אימון:

התקבל כשחקן ליגת כדורסל	שיתוף שחקני הקבוצה במשחק	רמת הבנת כלל המשחק	תדירות אימונים	מקום מגורים	ת.ז. שחקן
٥	ל	גבוהה	גבוהה	עיר	11111
٥	٥	גבוהה	גבוהה	עיר	22222
ל	ל	גבוהה	גבוהה	מושב	33333
ל	٥	גבוהה	בינונית	קיבוץ	44444
ל	٥	סבירה	נמוכה	קיבוץ	55555
٥	ל	סבירה	נמוכה	קיבוץ	66666
ל	٥	סבירה	נמוכה	מושב	77777
٥	ל	גבוהה	בינונית	עיר	88888
ל	ל	סבירה	נמוכה	עיר	99999
ל	٥	סבירה	בינונית	קיבוץ	10111
ל	٥	סבירה	בינונית	עיר	11222
ל	٥	גבוהה	בינונית	מושב	12333
ל	ל	סבירה	גבוהה	מושב	13444
٥	ל	גבוהה	בינונית	קיבוץ	14555

(המשך השאלה בעמוד הבא)

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, עבור נתוני האימון שבטבלה לחיזוי החלטה האם השחקן יתקבל לליגת כדורסל האוניברסיטה הפתוחה. בתשובתכם הדגימו את שלבי בחירת התכונה המפצלת בעץ.

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

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

#### שאלה 2 (clustering) – אשכול (15 נקודות)

בצעו אשכול היררכי בעזרת אלגוריתם מפלג (Divisive) לנתוני האימון שצוינו בשאלה 1.

#### בתשובתכם יש לכלול:

- אופן הגדרת המרחקים
- קריטריון לעצירת בניית הדנדרוגרם

#### שימו לב.

בתשובתכם הסופית יש להדגים את כל שלבי בניית הדנדרוגרם.

#### שאלה 3 (25 נקודות) – חוקי הקשר (association rules)

#### נתונה הטבלה:

מוצרים	מסי לקוח
מחשב, דיסק און קי, מדפסת, עכבר }	1
מחשב, דיסק און קי, מדפסת, אוזניות }	2
מחשב, דיסק און קי, אוזניות, כונן חיצוני, משחק מחשב}	3
{מחשב, מדפסת, משחק מחשב}	4

#### בהנחה:

Min support =50%

Min confidence = 60%

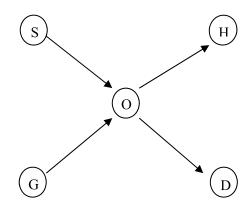
- א. מצאו את כל הקבוצות התדירות תוך שימוש באלגוריתם א-פריורי.
  - ב. הציגו את חוקי ההקשר החזקים.

### שאלה 4 (20 נקודות) –למידה בייאסינית

באיור הבא נתונה רשת בייסיאנית המתארת את העובדות הבאות:

השמנה בטנית (יסומן S) מתפתחת כתוצאה מאכילה מופרזת של סוכרים (יסומן או/ו כתוצאה מאכילה מתפתחת (G). מתורשה (יסומן G).

(D יסומן סוכרת למחלת (H יסומן) או $\prime$ ו למחלת ליתר לחץ ליתר לחץ ליתר (חץ דם יסומן).



P(S)=0.3 P(G)=0.4 : בהנחה ש

: נגדיר

 $P(O|S^G)=0.8$ 

 $P(O|S^{\bar{G}})=0.4$ 

 $P(O|S^{G})=0.6$ 

 $P(O|S^G)=0.2$ 

- P(O) א. חשבו את ההסתברות של נבדק ללקות ההסתברות של
- ב. חשבו את ההסתברות שנבדק בעל גורם תורשתי ילקה בהשמנה בטנית (P(O|G

המשך הבחינה בעמוד הבא

# חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

## שאלה 5 (10 נקודות)

בקורס כריית מידע הושמעה הטענה הבאה:

ייככל שגודל בסיס נתוני האימון והמבחן גדל, כך גדלה רמת הביטחון של הערכת ביצועי המסווגיי.

חוו דעתכם. הסבירו ונמקו.

## בהצלחה!