מספר התלמיד הנבחן

רשום את כל תשע הספרות

# האוניברסיטה

הדבק כאן את מדבקת הנבחן

מס' שאלון - 516

25

ביוני 2018

מס' מועד 82

סמסטר 2018ב

י"ב בתמוז תשע"ח

20595 / 4

שאלון בחינת גמר

20595 - כריית מידע

משך בחינה: שעות

> בשאלון זה 4 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 אלקטרוני שבאמצעותו ניתן לאצור מידע

לרבות מכשיר טלפון נייד, מחשב נישא, שעון חכם וכד'.

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.

שאלה 1 (30 נקודות) – עצי החלטה

נתונה טבלת נתוני אימון:

המשך	חדר כושר	בדיקת מאמץ	גובה שחקן	גיל שחקן	מס׳ סידורי
פעילות בליגה					
ל	כ	לא תקין	200	40	1
ל	ל	לא תקין	195	38	2
٥	٥	לא תקין	203	22	3
٥	٥	לא תקין	185	27	4
٥	٥	תקין	180	30	5
ל	ל	תקין	178	28	6
٥	ל	תקין	185	23	7
ל	٥	לא תקין	187	39	8
٥	٥	תקין	178	41	9
٥	٥	תקין	182	גשום	10
٥	ל	תקין	183	42	11
٥	ל	לא תקין	184	19	12
٥	٥	תקין	199	20	13
۲	ל	לא תקין	188	21	14

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, עבור נתוני האימון שבטבלה לחיזוי המשך פעילות שחקן בליגת כדורסל. 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 (20 נקודות) – ניתוח אשכולות

.k-means בצעו אשכול לחמש הנקודות הבאות תוך שימוש באלגוריתם

P1: (1,2,3)

P2: (0,1,2)

P3: (3,0,5)

P4: (4,1,3)

P5: (5,0,1)

#### בתשובתכם הניחו:

- א. יש לבצע אשכול לשני אשכולות בלבד.
- C2(0,1,1) C1(1,0,0) ב. מרכזי האשכולות בשלב ההתחלתי
  - ג. שימוש במרחק מנהטן.

שימו לב, בתשובתכם הסופית יש להדגים את כל השלבים וכן להגדיר את האשכולות הסופיים.

#### שאלה 3 (20 נקודות) – חוקי הקשר

נתונה הטבלה:

בוטנים	זיתים	ציפס	שרשרת	יין אדום	בירה	טרנזקציה
1	0	1	0	1	1	T1
0	1	1	1	1	1	T2
0	1	0	1	1	1	Т3
0	0	1	0	1	1	T4

הגדירו את כל חוקי ההקשר המקיימים:

Min support=60%

Min\_confidence=80%

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

בבדיקת צריכת סמים שעורכת משטרת הגבולות בשדה התעופה באמסטרדם ב-1% מהמקרים מתקבלת תוצאה חיובית שגויה - FP דהיינו 1% שאינם צורכים סמים תוצאת הבדיקה שלהם חיובית). כמו כן, 5% מהמקרים מתקבלת תוצאה שלילית שגויה - FN.

בהנחה ש-2% מהנבדקים צורכים סמים. חשבו מהי ההסתברות שנבדק שתוצאת הבדיקה שלו חיובית אכן באמת צורך סמים.

בתשובתכם, ציינו את כלל החישובים.

## חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

### שאלה 5 (10 נקודות) סיווג וחיזוי

במהלך קורס כריית מידע טען אחד הסטודנטים: יי בשיטת סיווג אין כל חשיבות לאופן חלוקת הנתונים לנתוני אימון ונתוני מבחן. יש רק להקפיד שנתוני המבחן יהיו בלתי תלויים בנתוני האימון".

חוו דעתכם.

בהצלחה!