מספר התלמיד הנבחן רשום את כל תשע הספרות

האוניברסיטה בפתוחה

999999999999999

הדבק כאן את מדבקת הנבחן

-

ביוני 2014

30

סמסטר 2014ב

ב' בתמוז תשע"ד

מס' מועד 82

מס' שאלון - 493

20595 / 4

שאלון בחינת גמר

20595 - כריית מידע

משך בחינה: 3 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.

השימוש במחשב נישא או במכשיר כלשהו שבאמצעותו אפשר להתחבר

לאינטרנט אסור.

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 (decision trees)

נתונה טבלת נתוני אימון:

סוג הכביש	תאונה	עקיפה	מזג אוויר	שעת הנהיגה	מס׳ נבדק
אוטוסטרדה	לא	לא	נאה	08: 00	1
אוטוסטרדה	כן	לא	מעונן	08:00	2
עירוני	לא	לא	נאה	10:00	3
אוטוסטרדה	לא	כן	גשום	09:00	4
אוטוסטרדה	כן	כן	נאה	09:00	5
עירוני	לא	לא	נאה	10:00	6
עירוני	לא	לא	מעונן	10:00	7
הררי	לא	לא	גשום	09:00	8
אוטוסטרדה	לא	כן	נאה	09:00	9
אוטוסטרדה	כן	כן	מעונן	10:00	10
עירוני	לא	לא	גשום	10:00	11
אוטוסטרדה	לא	כן	מעונן	08:00	12
הררי	לא	לא	נאה	09:00	13
אוטוסטרדה	לא	כן	מעונן	08:00	14

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד לנתוני האימון שבטבלה לחיזוי החלטה מהו סוג הכביש שבו נהג הנבדק. הדגימו בתשובתכם את שלבי בחירת התכונה המפצלת בעץ.

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

ב. איזו מבין התכונה/תכונות אפשר להסיר ומדוע! אם אין תכונה הניתנת להסרה, יש לציין זאת מפורשות.

### שאלה 2 (15 נקודות) – חוקי הקשר (association rules)

: min\_support את הנחת אפריורי אלגוריתם אפריורי שהוגדרו על ידי אלגוריתם אפריורי המקיים את מחונות 7 במונות abc, abc, abc, bcc, bcc, bcc, bdc, cde

בצעד הבא של האלגוריתם האפריורי, מי יהיו המועמדים הראשוניים של הקבוצות התדירות הבאות בהינתן שקבוצת הפריטים היא 4!

### שאלה 3 (20 נקודות) – ניתוח אשכולות (clustering)

. k-means בצעו אשכול לשבע הנקודות שלהלן תוך שימוש באלגוריתם

A1(0, 0), A2(8,0), A3(8,6)

B1(16,0), B2(0,6)

C1(16,6), C2(16,0)

בתשובתכם הניחו:

- א. יש לבצע אשכול לשלושה אשכולות בלבד.
- ב. מרכזי האשכולות בשלב ההתחלתי הם A1, B1, C1
- ג. שימוש במרחק אאוקלידי (Euclidian distance) בריבוע.

#### שימו לב,

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

#### שאלה 4 (25 נקודות) – למידה בייסיאנית

 $X_1; X_2, X_3$  בטבלה שלפניכם נתונים נתוני תצפיות בטבלה שלפניכם נתונים נתוני

Y	$X_3$	$X_2$	$X_1$
0	0	0	0
1	0	0	0
0	1	1	0
0	0	1	1
1	0	1	1
1	1	0	1
1	1	1	1

- א. מהן ההסתברויות החיוניות למסווג נאייב בייס (Naive Bayes)!
- $(X_1, X_2, X_3)$ =(0,0,0) עבור Y בייס לסיווג ערכו בייס בייס אווג ערכו בנאייב בייס לסיווג ערכו של

# חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

# שאלה 5 (10 נקודות)

לרעש יותר k-means במהלך קורס כריית מידע טען אחד הסטודנטים: "אלגוריתם מידע טען אחד חסין לרעש יותר מ-PAM". חוו דעתכם.

# בהצלחה!