

רשום את כל תשע הספרות

האוניברסיטה

י"ב בתמוז תשע"ה

מס' שאלון - 506

29

ביוני 2015

הדבק כאן את מדבקת הנבחן

סמסטר 2015ב

20595 / 4

מס' מועד 82

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 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.

נתונה טבלת נתוני אימון:

| טרשת | רמת | רמת | לחץ דם | לחץ דם | מעשן | מגדר | גיל | מסי |
|--------|--------------|----------|---------|---------|----------|------|-----|------|
| עורקים | טריגליצרידים | כולסטרול | דיסטולי | סיסטולי | | | | נבדק |
| ל | 80 | 125 | 75 | 120 | כ |) | 53 | 001 |
| X | 75 | 116 | 75 | 135 | う | 7 | 56 | 002 |
| ל | 82 | 130 | ?? | 126 | り | 7 | 45 | 003 |
| ٥ | 85 | 135 | 97 | 145 | ú | 7 | 44 | 004 |
| ל | 97 | 125 | 66 | 123 | ú | 7 | 58 | 005 |
| ٥ | 86 | 133 | 88 | 151 | ל | , | 60 | 006 |
| ٥ | 97 | 146 | 77 | ?? | ט | 7 | 43 | 007 |
| ל | 74 | 120 | 66 | 127 | ל | ۲ | 59 | 008 |
| ٥ | 92 | 132 | 88 | 145 | ל | ١ | 57 | 009 |

(data preprocessing) שאלה 1 (10 נקודות) – הכנת הנתונים

ציינו והדגימו את שלבי הכנת הנתונים לביצוע כריית מידע.

המשך הבחינה בעמוד הבא

שאלה 2 (30 נקודות) – עצי החלטה (decision tree)

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, עבור נתוני האימון שבטבלה לחיזוי מחלת טרשת עורקים. בתשובתכם הדגימו את שלבי בחירת התכונה המפצלת בעץ.

הערה: יש לכלול חישובים כדוגמת אנטרופיה, Gain ratio, מדד גיני.

רמז: יש להשתמש באחת משיטות החלוקה למרווחים (דיסקרטיזציה) הבאות:

- חלוקה לאינטרוולים בעלי רוחב זהה
- (bins) חלוקה לאינטרוולים בעלי תדירות זהה -
- ב. איזה מבין התכונה/תכונות ניתן להסיר ומדוע? אם אין תכונה הניתנת להסרה, יש לציין זאת מפורשות.

שאלה 3 (25 נקודות) – מסווג בייאסיני (Bayesian Classification)

סווגו האם לנבדק יש טרשת עורקים עבור מופע הבא <י', 271 י, 13, נ, ל, 124, 87,151, <97 תוך שימוש במסווג בייס.

בתשובתכם הדגימו את ההסתברויות המותנות אשר ערכתם בהן שימוש.

שאלה 4 (25 נקודות) – ניתוח אשכולות (clustering)

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

בתשובתכם יש לכלול:

- אופן הגדרת המרחקים
- קריטריון לעצירת בניית הדנדרוגרם

שימו לב,

בתשובתכם הסופית יש להדגים את כל שלבי בניית הדנדרוגרם.

המשך הבחינה בעמוד הבא

חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

שאלה 5 (10 נקודות)

במהלך קורס כריית מידע טען אחד הסטודנטים: ייבאלגוריתם אפריורי בכל מחזור יוצרים מועמדים מועמדים אשר אינם מקיימים את תכונת האפריורייי. חוו דעתכם.

בהצלחה!