מספר התלמיד הנבחן

רשום את כל תשע הספרות

האוניברסיטה

הדבק כאן את מדבקת הנבחן

14 מס' שאלון - 535

ביולי 2016

סמסטר 2016ב

ח' בתמוז תשע"ו

20595 / 4

85 מס' מועד

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 אלקטרוני שבאמצעותו ניתן לאצור מידע

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

החזירו

למשגיח את השאלון

וכל עזר אחר שקיבלתם בתוך מחברת התשובות

בהצלחה !!!

חלק א (90 נקודות)

בחלק זה ארבע שאלות. ענו במפורט על ארבע השאלות.

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

: נתונה טבלת נתוני אימון

היקף התקף הלב	לחץ דם גבוה מאוד	השמנה	רמת סוכר	מועד הופעת תסמיני תעוקה בחזה	מס' קריאה
גדול מאוד	לא	לא	תקין	בוקר	1
גדול מאוד	כן	לא	גבוה	בוקר	2
קטן	לא	לא	תקין	ערב	3
גדול מאוד	לא	כן	גבוה מאוד	צהריים	4
גדול מאוד	כן	כן	תקין	צהריים	5
קטן	לא	לא	תקין	ערב	6
קטן	לא	לא	גבוה	ערב	7
בינוני	לא	לא	גבוה מאוד	צהריים	8
גדול מאוד	לא	כן	תקין	צהריים	9
גדול מאוד	כן	כן	גבוה	ערב	10
קטן	לא	לא	גבוה מאוד	ערב	11
גדול מאוד	לא	כן	גבוה	בוקר	12
בינוני	לא	לא	תקין	צהריים	13
גדול מאוד	כן	כן	גבוה מאוד	צהריים	14

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, עבור נתוני האימון שבטבלה לחיזוי החלטה מהו היקף התקף הלב. בתשובתכם הדגימו את שלבי בחירת התכונה המפצלת בעץ.

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

ב. האם ניתן לגזום את עץ ההחלטה שבניתם בסעיף א! נמקו תשובתכם.

המשך הבחינה בעמוד הבא

(association rules) – חוקי הקשר – חוקי נקודות) – אאלה 2

נתונה הטבלה:

Items	TransID		
A, B, C, D	T100		
A, B, C, E	T200		
A, B, E, F,	T300		
Н			
A, C, H	T400		

: הניחו

Min_support = 50%

Min_confidence = 60%

חשבו את מדד הענין (lift) עבור כל הקבוצות התדירות.

שאלה 3 (clustering) - ניתוח אשכולות (clustering)

במפה שלפניכם מסומנות שש ערים באיטליה, עבורן בצד שמאל הוגדרה מטריצת המרחקים בקיימ בין הערים השונות.

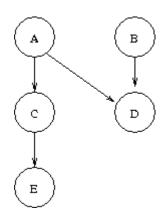
	BA	FI	МІ	NA	RM	TO
BA	0	662	877	255	412	996
FI	662	0	295	468	268	400
М	877	295	0	754	564	138
NA	255	468	754	0	219	869
RM	412	268	564	219	0	669
ТО	996	400	138	869	669	0



single- בצעו אשכול היררכי למרחקים בין הערים השונות באיטליה, תוך הסתייעות בשיטת . linkage

שאלה 4 (20 נקודות) – למידה בייאסינית

באיור הבא נתונה רשת בייסיאנית



.FALSE איין א - F - מציין איין דער א - TRUE להלן נתונות ההסתברויות, כאשר

Prob(A=T) = 0.3

Prob(B=T) = 0.6

Prob(C=T|A=T) = 0.8

Prob(C=T|A=F) = 0.4

Prob(D=T|A=T,B=T) = 0.7

Prob(D=T|A=T,B=F) = 0.8

Prob(D=T|A=F,B=T) = 0.1

Prob(D=T|A=F,B=F) = 0.2

Prob(E=T|C=T) = 0.7

Prob(E=T|C=F) = 0.2

חשבו את ההסתברויות הבאות:

Prob(D=T) א.

Prob(C=T | A=F, E=T) .ם

המשך הבחינה בעמוד הבא

חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

שאלה 5 (10 נקודות)

ומדד (sensitivity) הוא מדד הרגישות (accuracy) ומדד הדגימו כיצד מדד הדיוק (sensitivity) הוא פונקציה של מדד הרגישות (specificity) .

בהצלחה!