**ממ"ן 12**

**ניר פרלמן 031196744**



P(S = no|B = bad, F = not empty) =

1 – P(S = yes|B = bad, F = not empty) = 1 - 0.9 = 0.1

P(S = no|B = bad, F = empty) = 1 – P(S = yes|B = bad, F = empty) = 1 - 1 = 0

P(S = yes|B = bad) = P(S = yes|B = bad, F = not empty)\*P(F = not empty) +

P(S = yes|B = bad, F = empty)\*P(F = empty) = 0.1\*0.8 + 0\*0.2 = 0.08



P(B = good, F = empty, G = empty, S = yes) =

P(S = yes|B = good, F = empty)\*P(B = good)\*P(F = empty)\*

P(G = empty|B = good, F = empty) = 0.2\*0.9\*0.2\*0.8 = 0.0288



נסמן את חוגי הספורט כמספרים: כדורסל – 1, כדורגל – 2, שחמט – 3, פסנתר – 4,

דרמה – 5, אומנות – 6. נרשום את הטבלה הנתונה באמצעות המספרים שהגדרנו. בעמודה הימנית מספר התלמיד ובעמודה השמאלית מספרי החוגים עבור כל תלמיד.

|  |  |
| --- | --- |
| תלמיד | חוגים |
| 1 | 321 |
| 2 | 254 |
| 3 | 632 |
| 4 | 54 |
| 5 | 462 |
| 6 | 5432 |
| 7 | 164 |
| 8 | 31 |
| 9 | 5432 |
| 10 | 62 |

נבדוק תדירות עבור קבוצות של חוג אחד:

1. קיים 3 פעמים ולא עובר את אחוז התמיכה ולכן לא תדיר.
2. קיים 7 פעמים ולכן תדיר.
3. קיים 5 פעמים ולכן תדיר.
4. קיים 6 פעמים ולכן תדיר.
5. קיים 4 פעמים ולכן לא תדיר.
6. קיים 4 פעמים ולכן לא תדיר.

לפי אלגוריתם האפריורי, תת-קבוצות לא תדירות, יניבו קבוצות גדולות יותר לא תדירות. לכן לא נבדוק קבוצות המכילות תת-קבוצות שלהן שאינן תדירות.

קבוצה 2,3: קיימת רק 4 פעמים (צריך מינימום 5) ולכן לא תדירה.

קבוצה 2,4: קיימת רק 4 פעמים ולכן לא תדירה.

קבוצה 3,4: לא תדירה.

לכן לא מצאנו קבוצות תדירות בכלל. אין צורך בשימוש בביטחון כי אין חוקי הקשר.

1. ניקח את נתוני האמון מממ"ן 11:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Work as a guide | Teaching Experience | Work XP | WEKA? | Programming Language | Degree | Sex | Age | Id |
| no | no | no | yes | Java | Computer Science | f | 25 | 12345 |
| yes | yes | yes | no | C++ | Computer Science | f | 30 | 67891 |
| yes | no | yes | yes | Java | Management | m | 28 | 71294 |
| no | no | yes | yes | C++ | Computer Science | f | 35 | 34568 |
| yes | yes | yes | yes | C++ | Computer Science | f | 38 | 12578 |
| yes | yes | yes | yes | Java | Management | m | 33 | 25804 |
| no | no | yes | yes | C++ | Computer Science | m | 40 | 16923 |
| no | no | no | no | Java | Computer Science | f | 24 | 34890 |

נמחק את עמודת המין כי זה יחס סימטרי ואנחנו מחפשים מרחקים לא סימטריים.

ממוצע הגילאים הוא 32.25=8/(25+30+28+35+29+38+33+40)

נמיר את הגיל לבינארי לפי הממוצע: מעל ל32.25 ומתחת.

נמיר את שאר הערכים לבינארית בצורה הבאה:

1. No, java, computer science, age<32.25
2. Yes, c++, management, age>32.25

נקבל את הטבלה:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Work as a guide | Teaching Experience | Work XP | WEKA? | Programming Language | Degree | Age | Id | No. |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 12345 | 1 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 67891 | 2 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 71294 | 3 |
| 0 | 0 | 1 | 1 | 1 | 0 | 1 | 34568 | 4 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 12578 | 5 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 25804 | 6 |
| 0 | 0 | 1 | 1 | 1 | 0 | 1 | 16923 | 7 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 34890 | 8 |

נחשב את המרחקים בין כל הרשומות (נשתמש במספר סידורי במקום מספר הזהות):

נשתמש בחישוב אוקלידי לגילוי המרחק (טבלת אי הסימטריה) בין שתי רשומות.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |  |
|  |  |  |  |  |  |  | 0 | 1 |
|  |  |  |  |  |  | 0 | 2.236 | 2 |
|  |  |  |  |  | 0 | 2 | 1.73 | 3 |
|  |  |  |  | 0 | 2 | 2 | 1.73 | 4 |
|  |  |  | 0 | 1.414 | 2 | 1.414 | 2.236 | 5 |
|  |  | 0 | 1.414 | 2 | 1.414 | 2 | 2.236 | 6 |
|  | 0 | 2 | 1.414 | 0 | 2 | 2 | 1.73 | 7 |
| 0 | 2 | 2.449 | 2.449 | 2 | 2 | 2 | 1 | 8 |

לפי חישוב אשכול היררכי אגלומרטיבי של הדנדרוגרם נתחיל עם 8 אשכולות של רשומה אחת ונגדיל אותן לפי המרחקים הקרובים. הדבר מוצג בדיאגרמה הבאה:

0

1

1.414

1.73

0

1

2

3

שלב

מרחק

עוצרים את בניית הדנדרוגרם כאשר איחדתנו את כל האשכולות לאחד שמכיל את כל הרשומות.