

שאלה מס' 2 [60 נקודות]

חשוב: יש לפרט את כל החישובים וההנחות!

נתונות 8 טרנזקציות על הפריטים {A, B, C, D, E}:

tid	items
1	{A,B}
2	{A,B,C}
3	{B,C,D}
4	{B,C}
5	{A,B,C,D}
6	{B,D}
7	{B,E}
8	{B,C,D,E}

- א. יש למצוא את כל הצירופים השכיחים (frequent itemsets) תוך שימוש באלגוריתם Apriori בהינתן minimum support שווה ל- 35%. ציינו כל שלב של האלגוריתם. (20 נק)
- ב. יש לחשב את כל החוקים (Association Rules) מצירופים שמצאתם בסעיף א' בהינתן minimum confidence שווה ל- 70%. (10 נק)
- ג. נניח שאתם צריכים לבנות מודל מנבא לקניית פריט C בהינתן הטרנזקציות הנתונות של יתר הפריטים. נניח גם שעליכם להשתמש באלגוריתם ID3. באיזה אופן תציגו את הטרנזקציות לצורך הרצת האלגוריתם? (10 נק)
- ד. בהינתן הייצוג שהגדרתם בסעיף ג', תציגו את הרמה הראשונה של עץ ההחלטה המתקבל מהפעלת האלגוריתם ID3 המבוסס על Information Gain ללא pre-pruning. (20 נק)