1. בהינתן הטבלה הבאה:

|  |  |  |
| --- | --- | --- |
| קר | מעונן | גשום |
| כן | כן | כן |
| לא | כן | לא |
| לא | לא | לא |
| כן | לא | לא |
| לא | כן | כן |
| כן | לא | לא |
| לא | כן | לא |

1. חשבו support ו-confidence לחוקים: מעונן גורר גשום, מעונן וגם קר גורר גשום
2. הריצו ידנית את אלגוריתם אפריורי עבור min\_support = 0.3 ו- min\_confidence=0.8. פרטו את הריצה של האלגוריתם.
3. ממשו פונקציה support(df, rule) כאשר:
   1. df היא pandas data frame
   2. rule הוא מחרוזת מהצורה כאשר כל הוא שם של עמודה ב-data frame.
   3. הפונקציה מחזירה את ה- support של החוק ב-df.
4. ממשו פונקציות עם חתימה דומה לשאלה 2 שמממשות confidence ו-lift.
5. ציינו (עם דוגמאות) בעיות עיקריות של המדדים השונים להערכת חוקים.
6. כמה חוקים אפשריים קיימים על מידע עם n פיצ'רים כאשר לכל פיצ'ר יש m ערכים אפשריים?
7. ממשו פרוצדורה apriori(df, min\_conf, min\_support) שמקבלת כקלט טבלה וערכי סף ל-confidence ו-support. הפונקציה תחזיר את כל החוקים עם support ו-confidence הגבוהים מערכי הסף הנ"ל (בפורמט זהה לזה שבשאלה 2).
8. שאלה זו מנסה להדגים ש-confidence גבוה לא אומר בהכרח שהחוק המוחזר אכן שריר במציאות (כלומר הסתכלות על confidence גבוה לא חשוב רק בשביל למצוא חוקים משמעותיים שקיימים הרבה אלא כדי לנפות חוקים הנובעים מאקראיות גרידא):

נניח שבמידע שלנו יש m פיצ'רים אקראיים לגמרי (רעש). הראו שעבור m מספיק גדול כמעט בוודאות יהיה חוק עם confidence 1.