2. The following contingency table summarizes supermarket transaction data. (It is similar in format to the table you see on slide 20 in our lecture notes on association rules.  means sum by row or column.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Hot dog | | No hot dog | | ∑ | |
| Hamburger | | 1800 | | 700 | | 2500 |
| No hamburger | | 1200 | | 1300 | | 2500 |
| ∑ | | 3000 | | 2000 | | 5000 |

1. Suppose that the association rule "hot dogs  hamburgers" is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule valid? **(0.5 Point)**

The given relation is

HOT Dog => HAMBURGER

Confidence for the above relation is = # HOT DOGS AND HAMBURGER / # HOT DOGS

= 1800/3000

= 0.6 OR 60%

Support for the same equation is = # HOT DOGS AND HAMBURGERS / # OF ITEMS TOTAL

= 1800/5000

= 0.36 OR 36%

GIVEN:

Minimum Support < Support

Minimum Confidence < Confidence

ALSO,

LIFT = Confidence of the rule/ Support of the HAMBURGER

= 0.6 / (1/2)

= 1.2 > 1

Thus, the above said rule is a VALID RULE.

1. Based on the given data, is the purchase of hot dogs independent of the purchase of hamburgers? If not, what kind of correlation relationship exists between the two (i.e., if a customer purchases hot dogs, will that increase or decrease her chance of purchasing hamburgers)? (**1 Point)**

Given:

HOT DOG => HAMBURGER

SUPPORT = 36%

CONFIDENCE = 60%

HAMBURGER => HOT DOG

SUPPORT = 700/5000 = 14%

CONFIDENCE = 700/2500 = 28%

Thus,

The probability that someone who has bought HOT DOG will also buy HAMBURGER is 60%

The probability that someone who has bought HAMBURGER will also buy HOT DOG is 28%

Also, P(HOT DOG ^ HAMBURGER) = 1800/5000 = 36 %

& P (HOT DOG) \* P(HAMBURGER) = (3000/5000) \* (2500/5000) = 30%

Since, the probability are not equal. It means they are not independent.