Question-1:

Firstly, I will calculate the support counts for all individual items.

- ⇨ **M = {Beef: 3/7 ≈ 43%, Boots: 1/7 ≈ 14%, Cheese: 3/7 ≈ 43%, Clothes: 3/7 ≈ 43%, Shoes: 1/7 ≈ 14%, Milk: 3/7 ≈ 43%, Bread: 6/7 ≈ 86%}**

So, ‘Boots’ and ‘Shoes’ will not go to L because they do not satisfy MIS value of first item.

- ⇨ **L = {Beef, Cheese, Clothes, Milk, Bread}**

Now, I must check satisfying own MIS values of items in L for going to set of F_1. All items satisfy own MIS values except ‘Milk’, because MIS value of Milk is 50% but support count of Milk is 43%.

- ⇨ **F_1 = {Beef, Cheese, Clothes, Bread}** // The information about F_1 is given in the question but I must know itemset which is called ‘L’ to find the set ‘C_2’.

<u>To find set of C_2,</u>

Statements of for-loop are satisfying based on 4th line in Algorithm MS-Apriori with ‘k = 2’ and ‘$F_1 \neq \varnothing$’. Then, if-statement is satisfied in 5th line with ‘k = 2’, too. So, I must go to Function level2-candidate-gen. I am creating an empty set which is called ‘C_2’. Then, I am taking the first item in L -Beef- and check satisfying own MIS value. It satisfies and I am taking the item that is after ‘Beef’, it is ‘Cheese’. I check satisfying MIS value of ‘Beef’ and Bingo! This pair of items fulfil the all conditions and deserve being incorporated into ‘C_2’. I will check all situations about deserving being incorporated into ‘C_2’ for all pairs and I will calculate the itemset ‘C_2’. So,

C_2 = {{Beef, Cheese}, {Beef, Clothes}, {Beef, Milk}, {Beef, Bread}, {Cheese, Clothes}, {Cheese, Milk}, {Cheese, Bread}, {Clothes, Milk}, {Clothes, Bread}}

<u>To find set of F_2,</u>

I am going back to 9th line of Algorithm MS-Apriori and now, I am taking the first transaction -Beef, Bread-. According to for-loop in 10th line, I am taking the first candidate in C_2 -Beef, Cheese- and I am checking this candidate is contained in this transaction. If the answer is yes, I will add to support count of this candidate. This process is going on for all candidates in C_2. When all candidates are checked, I will go back to first for-loop and I will take the second transaction. Then, I will check all candidates until all transactions are checked.

C_2 = {{Beef, Cheese}: 2/7 ≈ 29%, {Beef, Clothes}: 0/7 = 0%, {Beef, Milk}: 1/7 ≈ 14%, {Beef, Bread}: 3/7 ≈ 43%, {Cheese, Clothes}: 0/7 = 0%, {Cheese, Milk}: 1/7 ≈ 14%, {Cheese, Bread}: 2/7 ≈ 29%, {Clothes, Milk}: 2/7 ≈ 29%, {Clothes, Bread}: 3/7 ≈ 43%}

Finally, a candidate which satisfies MIS value of first item of this candidate is adding to itemset ‘F_2’. So,

F_2 = {{Beef, Cheese}, {Beef, Bread}, {Cheese, Bread}, {Clothes, Milk}, {Clothes, Bread}}

To find set of C_3,

I am going back to 4th line in Algorithm MS-Apriori. Statements of for-loop are satisfying with 'k = 3' and '$F_2 \neq \emptyset$'. Then, , if-statement is not satisfied in 5th line and I am going to 7th line which defines the C_3 using Function MScandidate-gen. So, I am going to MScandidate-gen function. First of all, I am creating the candidate itemset which is called 'C_3'. Now, I am searching for f_1 and f_2 which are same except last item of them. Also, last item of f_2 must be greater than last item of f_1 according to MIS value and if MIS values are equal, I must check lexicographic order. So, I found two pair.

1. The first of these pairs,

- **f_1: Beef, Bread**
- **f_2: Beef, Cheese**

I am identifying a candidate which is named 'c' and so, **c = {Beef, Bread, Cheese}.** I am adding this 'c' to the C_3 and **C_3 = {{Beef, Bread, Cheese}}** for now. In 8th and 9th lines, these if-statements are for removing c from C_3. For each all 2-subsets which is named 's_1, s_2 and s_3'.

1. **s_1: Beef, Bread**
   First if-statement is true, because first item of c is in s_1. Second if-statement is not true because s_1 is in F_2. So, it will not be deleted for now.
2. **s_2: Beef, Cheese**
   First if-statement is true, because first item of c is in s_2. Second if-statement is not true because s_2 is in F_2. So, it will not be deleted for now.
3. **s_3: Bread, Cheese**
   First if-statement is not true, because first item of c is not in s_3 OR MIS values of Bread and Cheese are not equal. So, c is not deleted from C_3.

2. The second of these pairs,

- **f_1: Clothes, Bread**
- **f_2: Clothes, Milk**

I am identifying a candidate which is named 'c' and so, **c = {Clothes, Bread, Milk}.** I am adding this 'c' to the C_3 and **C_3 = {{Beef, Bread, Cheese}, {Clothes, Bread, Milk}}** for now. In 8th and 9th lines, these if-statements are for removing c from C_3. For each all 2-subsets which is named 's_1, s_2 and s_3'.

1. **s_1: Clothes, Bread**
   First if-statement is true, because first item of c is in s_1. Second if-statement is not true because s_1 is in F_2. So, it will not be deleted for now.
2. **s_2: Clothes, Milk**
   First if-statement is true, because first item of c is in s_2. Second if-statement is not true because s_2 is in F_2. So, it will not be deleted for now.
3. **s_3: Bread, Milk**
   First if-statement is not true, because first item of c is not in s_3 OR MIS values of Bread and Cheese are not equal. So, c is not deleted from C_3.

Finally, I found the C_3 and it is,

C_3 = {{Beef, Bread, Cheese}, {Clothes, Bread, Milk}}

To find set of F_3,

I am going back to 9th line of Algorithm MS-Apriori and now, I am taking the first transaction -Beef, Bread-. According to for-loop in 10th line, I am taking the first candidate in C_3 -Beef, Bread, Cheese- and I am checking this candidate is contained in this transaction. If the answer is yes, I will add to support count of this candidate. This process is going on for all candidates in C_3. When all candidates are checked, I will go back to first for-loop and I will take the second transaction. Then, I will check all candidates until all transactions are checked.

C_3 = {{Beef, Bread, Cheese}: 2/7 ≈ 29%, {Clothes, Bread, Milk}: 2/7 ≈ 29%}

Finally, a candidate which satisfies MIS value of first item of this candidate is added to itemset 'F_3'. So,

F_3 = {{Beef, Bread, Cheese}, {Clothes, Bread, Milk}}

Question-2:

I am calculating the values of probability of Class is T and Class is F.

$\mathbf{Pr}(\boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{5}}{\mathbf{10}} = \frac{\mathbf{1}}{\mathbf{2}}$	$\mathbf{Pr}(\boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{5}}{\mathbf{10}} = \frac{\mathbf{1}}{\mathbf{2}}$

I am calculating the values of probability about variables of 'Attr1' when Class is T, and Class is F.

$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x1} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{2}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x1} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{1}}{\mathbf{5}}$
$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x2} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{2}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x2} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{2}}{\mathbf{5}}$
$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x3} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{1}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr1} = \boldsymbol{x3} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{2}}{\mathbf{5}}$

I am calculating the values of probability about variables of 'Attr2' when Class is T, and Class is F.

$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x4} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{1}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x4} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{2}}{\mathbf{5}}$
$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x5} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{2}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x5} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{1}}{\mathbf{5}}$
$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x6} \mid \boldsymbol{Class} = \boldsymbol{T}) = \frac{\mathbf{2}}{\mathbf{5}}$	$\mathbf{Pr}(\boldsymbol{Attr2} = \boldsymbol{x6} \mid \boldsymbol{Class} = \boldsymbol{F}) = \frac{\mathbf{2}}{\mathbf{5}}$

1. Attr1 = x1 and Attr2 = x6, so Class = ?

$$\Pr(Class = T) \times \Pr(Attr1 = x1 \mid Class = T) \times \Pr(Attr2 = x6 \mid Class = T) = \frac{1}{2} \times \frac{2}{5} \times \frac{2}{5} = \frac{4}{50}$$

$$\Pr(Class = F) \times \Pr(Attr1 = x1 \mid Class = F) \times \Pr(Attr2 = x6 \mid Class = F) = \frac{1}{2} \times \frac{1}{5} \times \frac{2}{5} = \frac{2}{50}$$

So,

probability of Class is T is greater than probability of Class is F. Consequently, **Class must be T.**

2. Attr1 = x2 and Attr2 = x4, so Class = ?

$$\Pr(Class = T) \times \Pr(Attr1 = x2 \mid Class = T) \times \Pr(Attr2 = x4 \mid Class = T) = \frac{1}{2} \times \frac{2}{5} \times \frac{1}{5} = \frac{2}{50}$$

$$\Pr(Class = F) \times \Pr(Attr1 = x2 \mid Class = F) \times \Pr(Attr2 = x4 \mid Class = F) = \frac{1}{2} \times \frac{2}{5} \times \frac{2}{5} = \frac{4}{50}$$

So,

probability of Class is F is greater than probability of Class is T. Consequently, **Class must be F.**