Online Test 3: Association Rules

Test und Assessment – Druckansicht

Online Test 3: Association Rules

Datum: Sun Mar 13 14:13:13 2022 Maximale Punktezahl: 11

Frage 1 - Itemsets - Frequent, Maximal and Closed (1 Punkt) [ID: 1051797]

Let the following transactions be given:

ID	Items
1	A, B
2	A
3	A, B, C

Evaluate the following statements, assuming a support threshold of $\frac{2}{3}$:

Für jede Aussage muss entschieden werden: [true] oder [false]

true	false	
○	◉	A is an itemset.
◉	○	{A} is frequent.
○	◉	{A} is maximal.
◉	○	{A} is closed.

A is not an itemset, but a single item.

{A} passes the support threshold and thus is a frequent itemset.

As there is no superset with the same support, {A} is closed as well.

However, {A} is not maximal, as {A, B} is a frequent superset.

Frage 2 - Association Rules - Metrics (1 Punkt) [ID: 1068063]

Let the following transactions be given:

ID	Items
1	fruits, milk, vegetables
2	soda, vegetables, milk, fruits
3	vegetables, fruits, soda, beer, milk
4	fruits, soda, milk
5	fruits, beer, milk
6	fruits, vegetables, soda, milk, beer
7	soda, milk, beer, vegetables, fruits
8	soda, fruits

The rule $\{vegetables\} \rightarrow \{milk\}$ has a (relative, i.e., in the range $[0, 1]$) support of [0.625] *(0.34 Punkte)* .

The confidence of this rule is [1] *(0.33 Punkte)* .

The lift of the rule is [1.143] *(0.33 Punkte)*

(Round your solutions to three decimal places if necessary and use a point (.) as decimal separator.)

$Support(\{vegetables\} \rightarrow \{milk\}) = P(\{vegetables, milk\}) = \frac{5}{8} = 0.625$

$Confidence(\{vegetables\} \rightarrow \{milk\}) = \frac{Support(\{vegetables\} \rightarrow \{milk\})}{Support(\{vegetables\})} = \frac{\frac{5}{8}}{\frac{5}{8}} = 1$

$Lift(\{vegetables\} \rightarrow \{milk\}) = \frac{Support(\{vegetables\} \rightarrow \{milk\})}{Support(\{vegetables\}) \cdot Support(\{milk\})} = \frac{\frac{5}{8}}{\frac{5}{8} \cdot \frac{7}{8}} = 1.143$

Frage 3 - Association Rules - Support (1 Punkt) [ID: 1051798]

Let the following transactions be given:

ID	Items
1	fruits, milk, vegetables
2	soda, vegetables, milk, fruits
3	vegetables, fruits, soda, beer, milk
4	fruits, soda, milk
5	fruits, beer, milk
6	fruits, vegetables, soda, milk, beer
7	soda, milk, beer, vegetables, fruits
8	soda, fruits

Which of the following association rules pass a support threshold of 40%?

Für jede Aussage muss entschieden werden: [yes] oder [no]

yes	no	
◉	○	$\{fruits\} \rightarrow \{milk\}$
○	◉	$\{soda\} \rightarrow \{beer\}$
◉	○	$\{vegetables\} \rightarrow \{milk\}$
◉	○	$\{milk, soda\} \rightarrow \{vegetables\}$

$Support(LHS \rightarrow RHS) = Support(LHS \cup RHS)$

This means support values are as follows:

$Support(\{fruits\} \rightarrow \{milk\}) = Support(\{fruits, milk\}) = \frac{7}{8}$

$Support(\{soda\} \rightarrow \{beer\}) = Support(\{soda, beer\}) = \frac{3}{8}$

$Support(\{vegetables\} \rightarrow \{milk\}) = Support(\{vegetables, milk\}) = \frac{5}{8}$

$Support(\{milk, soda\} \rightarrow \{vegetables\}) = Support(\{milk, soda, vegetables\}) = \frac{4}{8}$

Frage 4 - Association Rules - Confidence (1 Punkt) [ID: 1051799]

Let the following transactions be given:

ID	Items
1	fruits, milk, vegetables
2	soda, vegetables, milk, fruits
3	vegetables, fruits, soda, beer, milk
4	fruits, soda, milk
5	fruits, beer, milk
6	fruits, vegetables, soda, milk, beer
7	soda, milk, beer, vegetables, fruits
8	soda, fruits

Which of the following association rules pass a confidence threshold of 90%?

Für jede Aussage muss entschieden werden: [yes] oder [no]

yes	no	
○	◉	$\{fruits\} \rightarrow \{milk\}$
○	◉	$\{soda\} \rightarrow \{beer\}$
◉	○	$\{vegetables\} \rightarrow \{milk\}$
○	◉	$\{milk, soda\} \rightarrow \{vegetables\}$

$Confidence(LHS \rightarrow RHS) = \frac{Support(LHS \rightarrow RHS)}{Support(LHS)}$

This means confidence values are as follows:

$Confidence(\{fruits\} \rightarrow \{milk\}) = \frac{7}{8}$

$Confidence(\{soda\} \rightarrow \{beer\}) = \frac{3}{6}$

$Confidence(\{vegetables\} \rightarrow \{milk\}) = \frac{5}{5}$

$Confidence(\{milk, soda\} \rightarrow \{vegetables\}) = \frac{4}{5}$

Frage 5 - Apriori - Steps (1 Punkt) [ID: 1051800]

Order the steps of the Apriori algorithm!

Find frequent 1-itemsets

[BEGIN] Loop over k

Join

Prune

Count support

[END] Loop over k

Generate association rules

See the lecture.

Frage 6 - Apriori - Anti-Monotonicity (1 Punkt) [ID: 1068066]

Let $\{A, B, C\}$ be an itemset. Which of the following statements are true?

- ☐ If $\{A, B\}$ is frequent, then $\{A, B, C\}$ is frequent.
  (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.25 Punkte)
- ☑ If $\{A, B\}$ is not frequent, then $\{A, B, C\}$ is not frequent.
  (Ausgewählt = 0.25 Punkte, Nicht ausgewählt = 0 Punkte)
- ☑ If $\{A, B, C\}$ is frequent, then $\{A, B\}$ is frequent.
  (Ausgewählt = 0.25 Punkte, Nicht ausgewählt = 0 Punkte)
- ☐ If $\{A, B, C\}$ is not frequent, then $\{A, B\}$ is not frequent.
  (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.25 Punkte)

Anti-monotonicity means:

- Any nonempty subset of a frequent itemset is frequent too.
- Any superset of non-frequent itemset is non-frequent too.

Frage 7 - Apriori - Pruning (1 Punkt) [ID: 1051801]

Let the following frequent itemsets of size two be given (there are no further frequent itemsets of size two):

{A, B}

{A, C}

{A, D}

{B, D}

{C, D}

Which of the following itemsets of size three can be pruned in the Apriori algorithm?

Für jede Aussage muss entschieden werden: [yes] oder [no]

yes	no	
◉	○	{A, B, C}
○	◉	{A, B, D}
○	◉	{A, C, D}
◉	○	{B, C, D}

Apriori can prune all itemsets of size three that contain a non-frequent subset of size two. For each itemset of size three, there are three such subsets to be checked. Here, the itemset {B, C} is not frequent and can be used for pruning.

Frage 8 - Apriori Improvements - Hash Filter (1 Punkt) [ID: 1068068]

Let the following transactions be given:

{1, 2, 3}

{3, 5}

{1, 2, 4}

{2}

{1, 3}

Assume we have a hash filter with the function $(\sum_i a_i)\ mod\ 3$. Which of the following 2-itemsets can be pruned when scanning the database for 1-itemsets, assuming a support threshold of 60%?

- ☐ {1,2} *(Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.33 Punkte)*
- ☑ {1,3} *(Ausgewählt = 0.34 Punkte, Nicht ausgewählt = 0 Punkte)*
- ☐ {2,3} *(Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.33 Punkte)*

Initialization of hash filter: 0|0|0

After 1st transaction (contains three 2-itemsets): 1|1|1

After 2nd transaction (contains one 2-itemset): 1|1|2

After 3rd transaction (contains three 2-itemsets): 3|1|3

After 4th transaction (contains no 2-itemsets): 3|1|3

After 5th transaction (contains one 2-itemset): 3|2|3

The hash values of the 2-itemsets mentioned in the task are:

$h(\{1, 2\}) = 0$

$h(\{1, 3\}) = 1$

$h(\{2, 3\}) = 2$

Itemset {1, 3}'s hash value does not meet the support of 60%, so this itemset can be pruned. Note that ac-

tually all 2-itemsets do not meet the support threshold, but not all of them can be pruned, due to hash collisions.

Frage 9 - FP-Tree - Sorting (1 Punkt) [ID: 1051802]

Let the following transactions be given. Order the items according to their order in an FP-tree (assuming the items form a path in the tree and the root node is at the top)!

ID	Items
1	fruits, milk, vegetables
2	soda, vegetables, milk, fruits
3	vegetables, fruits, soda, beer, milk
4	fruits, soda, milk
5	fruits, beer, milk
6	fruits, vegetables, soda, milk, beer
7	soda, milk, beer, vegetables, fruits
8	soda, fruits

fruits

milk

soda

vegetables

beer

The items are ordered according to their frequency over all transactions (fruits: 8, milk: 7, soda: 6, vegetables: 5, beer: 4). More frequent items appear closer to the root of the tree.

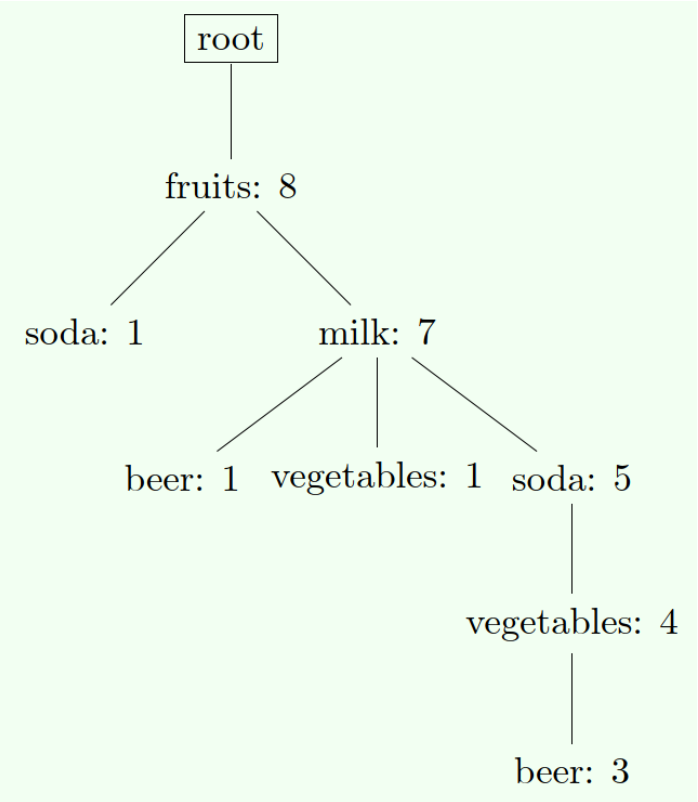
Frage 10 - FP-Tree - Construction (1 Punkt) [ID: 1051803]

Let the following transactions be given. How many edges does the corresponding FP-tree contain (the first edge starts from a dedicated "root" node, the remaining nodes correspond to the items)?

ID	Items
1	fruits, milk, vegetables
2	soda, vegetables, milk, fruits
3	vegetables, fruits, soda, beer, milk
4	fruits, soda, milk
5	fruits, beer, milk
6	fruits, vegetables, soda, milk, beer
7	soda, milk, beer, vegetables, fruits
8	soda, fruits

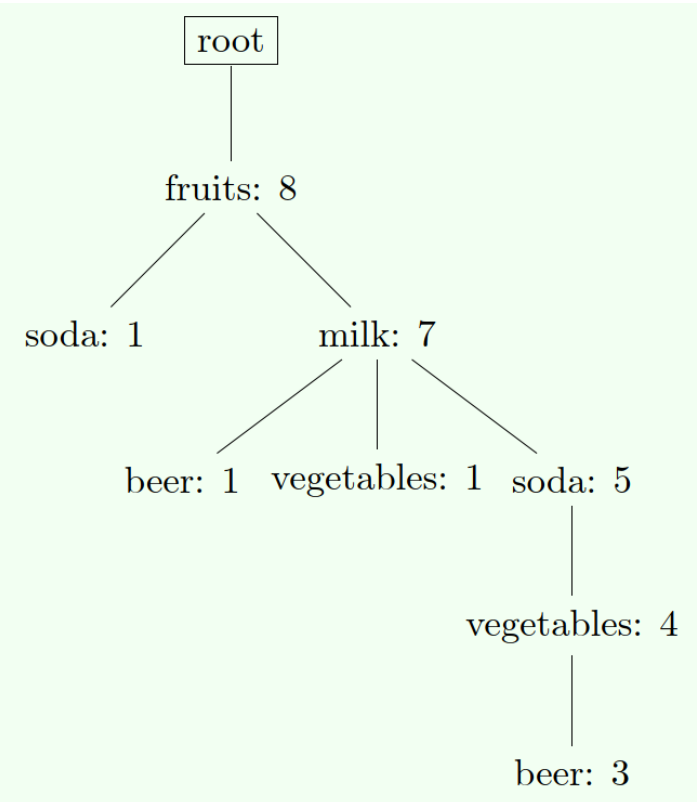
Der Wert muss zwischen 8 und 8 liegen

An exemplary tree could look like this (the horizontal order of child nodes is arbitrary here):



Frage 11 - FP-Tree - Rule Extraction (1 Punkt) [ID: 1051804]

Let the following FP-tree be given:



Which items are part of at least one frequent itemset also containing *vegetables*, assuming the absolute support threshold is five?

Für jede Aussage muss entschieden werden: [yes] oder [no]

yes	no	○	◉
beer		◉	○
fruits		◉	○
milk		○	◉
soda		There are two paths containing *vegetables*. The path via *fruits* and *milk* has a support of one. The path via *fruits*, *milk* and *soda* has a support of four till *vegetables* and a support of three till *beer*. As a result, *vegetables* occur five times together with *fruits* and *milk*, four times with *soda* and three times with *beer*.	