

Data Science & Machine Learning for engineering applications

AA 2023-2024

DataBase and Data Mining Group

Prof. Tania Cerquitelli







Given the confusion matrix depicted in figure, which statement **is not** correct?

Question 1

		Predicted		
		T	F	
Actual	Т	90	0	
	F	10	0	

- (a) Precision of class F is 0
- (b) Precision of class T is 90%
- $\bigcirc\,$ (c) All 100 items are predicted as class T
- (d) Model accuracy is 90%
- (e) Recall of class F is 10%
- (f) Recall of class T is 100%







Given the confusion matrix depicted in figure, which statement is not correct?

Question 2

		Predicted	
		T	F
Actual	Т	90	10
	F	0	0

- (a) Recall of class F is 0
- (b) Precision of class T is 100%
- (c) All elements belonging to class T are correctly predicted
- (d) Recall of class T is 90%
- (e) Precision of class F is 0
- (f) The model accuracy is 90%







Which of the following techniques can be used to avoid the "curse of dimensionality"?

- (a) None of the other answers is correct
- (b) Outlier Detection
- (c) Discretization
- (d) Stratified Sampling
- (e) Normalization
- (f) Feature Selection







Which of the following techniques can be used to reduce the number of records in a dataset?

- (b) None of the other answers is correct
- (c) Outlier Detection
- (d) Stratified Sampling
- (e) Normalization
- (f) Principal Component Analysis







A binary decision tree is used as a classification algorithm on a dataset containing 1000 records, 900 belonging to class A and 100 belonging to class B. Which of the following metrics allows an appropriate evaluation of the classifier's performance with regard to class B predictions?

- (a) Accuracy
- (b) Gini Index
- (c) Sum of Squared Error (SSE)
- (d) Mean Squared Error (MSE)
- (e) F-measure
- (f) Average Silhouette Index







Which of the following metrics **is not** suitable for evaluating the performance of a classification algorithm?

- (a) Precision
- (b) Confusion matrix
- (c) Silhouette Index
- (d) Accuracy
- (e) F-measure
- (f) Recall







The following transactional dataset is given.

ВD

ACDE

ВD

ΑE

BCDE

ABCD

AΒ

ABCE

BCD

ABCE

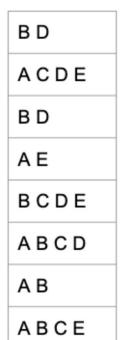
Write the header table of the corresponding FP-Tree with MinSup>2.







The following transactional dataset is given.



BCD

ABCE

Write the FP-Tree with MinSup>2. Specifically, report the list of paths characterizing the FP-Tree. For each path specify the sequence of nodes in the form (item, local support).







The following transactional dataset is given.

ВD

ACDE

ВD

ΑE

BCDE

ABCD

ΑВ

ABCE

BCD

ABCE

Write the node link chain for item E







The Apriori principle states that, if an itemset is frequent, then all of its subsets must also be frequent.

You are given the following transactional dataset.

rou are given the
ACE
ABCD
CD
ACD
АВ
ABE
BCE
ADE
ABE
ACDE

You need to apply the Apriori algorithm to extract the frequent itemsets with minsup = 2 (an itemset is frequent if it is contained in at least 2 transactions).

During the generation of the candidate itemsets of length 3, after the "prune step" with the application of the "Apriori Principle", what is the number of candidates extracted for counting their support in the dataset?



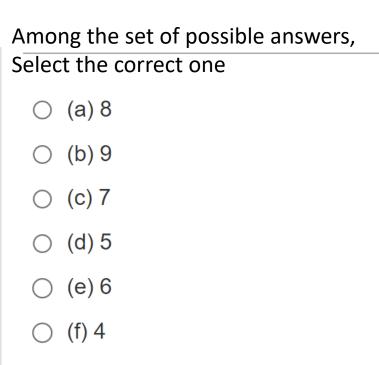




The Apriori principle states that, if an itemset is frequent, then all of its subsets must also be frequent.

You are given the following transactional dataset.

ACE
ABCD
CD
ACD
АВ
ABE
BCE
ADE
ABE
ACDE



You need to apply the Apriori algorithm to extract the frequent itemsets with minsup = 2 (an itemset is frequent if it is contained in at least 2 transactions).

During the generation of the candidate itemsets of length 3, after the "prune step" with the application of the "Apriori Principle", what is the number of candidates extracted for counting their support in the dataset?