

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

Question 1 of 20 > >>

Question 1

5 points Save Answer

A telecommunication company develops a prediction model to predict whether or not a customer is going to churn and plans to launch a customer care program to maintain those customers who are predicted to churn. In order to test the model's performance, the following confusion matrix is generated.

		Predicted Class	
		Churn=Yes	Churn=No
Actual Class	Churn=Yes	250	50
	Churn=No	100	600

Suppose this company has calculated the cost for each prediction as below.

		Predicted Class	
		Churn=Yes	Churn=No
Actual Class	Churn=Yes	\$4	\$67
	Churn=No	\$31	\$-7

Please compute the total cost of this prediction model (just type the number without dollar signs as your answer such as 12345).

→ ⚠ Moving to another question will save this response.

Question 1 of 20 > >>

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌵ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 2 of 20 ⏪ ⏩

Question 2

4 points  Saved

Please compute the Euclidean distance for the two data points $(-1, 8)$ and $(-16, 26)$. Round your answer to the second decimal place (e.g., 12.34). You may need to use calculator or Excel to assist you in this question.

23.43

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 2 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.
You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

✖ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 3 of 20 ⏪ ⏩

Question 3

4 points ✓ Saved

Based on the Apriori principle, if itemset AD is considered as frequent, which of the following is true?

- ☒ A. Its subsets such as A must be frequent.
- ☐ B. Its supersets such as A must be frequent.
- ☐ C. Its subsets such as ABD must be frequent.
- ☐ D. Its subsets such as A may be infrequent.
- ☐ E. Its supersets such as ABD must be infrequent.
- ☐ F. Its supersets such as ABD must be frequent.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 3 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

✖ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 4 of 20 ⏪ ⏩

Question 4

3 points  Saved

Which of the following is NOT a good practice to maintain ethical Data Mining?

- ☐ A. Anonymizing observations through removal of names and personally identifiable information.
- ☒ B. Collecting consumers' data such as gender, age, and height without their consent.
- ☐ C. Storing customer data in secure and protected environments.
- ☐ D. Respecting for individual rights.
- ☐ E. Providing clear and transparent information about data collection.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 4 of 20 ⏪ ⏩

Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 5 of 20 > ⏩

Question 5

5 points  Saved

Which of the following is TRUE of text mining and data mining? Choose all that apply.

- ☐ A. Text mining may involve a lot of noisy data such as spelling mistakes, but data mining never involves noisy data.
- ☒ B. The dataset in data mining usually has more attributes than that in text mining.
- ☒ C. Some data mining methods such as clustering or classification can be used in text mining.
- ☐ D. Text mining is a simple and straightforward process, requiring minimal effort and resources.
- ☐ E. Text mining was broadly implemented earlier than data mining.
- ☒ F. The data in text mining is usually semi-structured or unstructured, but the data in data mining is usually structured.
- ☒ G. Text mining usually deals with textual data, while the data mining deals with numerical and categorical data.

→ ⚠ Moving to another question will save this response.

⏪ < Question 5 of 20 > ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌵ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 6 of 20 ⏪ ⏩

Question 6

3 points  Saved

Which of the following about hierarchical clustering is NOT correct?

- ☐ A. It organizes a set of nested clusters as a hierarchical tree
- ☒ B. It will always create a specific number of clusters
- ☐ C. It is sensitive to outliers
- ☐ D. It does not work the best when you have a large amount of data (e.g., thousands or millions).
- ☐ E. It includes two main types, agglomerative and divisive.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 6 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 7 of 20 ⏪ ⏩

Question 7

4 points

Save Answer

Please compute the Manhattan distance between the two data points (15.70,-14.75) and (-5.09,11.78). Round your answer to the second decimal place (e.g., 12.34). You may need to use calculator or Excel to assist you in this question.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 7 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 8 of 20 > ⏩

Question 8

3 points ✓ Saved

Which of the following about association rule mining is correct?

- ☐ A. Based on Apriori principle, if the rule {Bread} → {Milk} has a support of 10%, then the support of {Bread, Beer} → {Milk} should be no greater than 10%.
- ☒ B. A rule's support value is never equal to its confidence value.
- ☐ C. A rule's support value is always greater than its confidence value.
- ☐ D. The rule {Bread} → {Milk} and the rule {Milk} → {Bread} are supposed to have the same confidence value.
- ☐ E. A strong rule means the rule has a very high confidence value, regardless of its support value.
- ☐ F. Because actionable rules contain high-quality and actionable information, they occur most often among the three types of association rules.

→ ⚠ Moving to another question will save this response.

⏪ < Question 8 of 20 > ⏩

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

✖ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 9 of 20 ⏪ ⏩

Question 9

3 points ✓ Saved

How many bigrams and trigrams are in the following sentence?

"In the realm of technology, artificial intelligence such as Gemini is advancing rapidly, fundamentally altering the way we interact with machine".

Please note that you do not need to remove stop words in your count.

- ☐ A. 19 bigrams and 20 trigrams
- ☐ B. 18 bigrams and 19 trigrams
- ☒ C. 20 bigrams and 19 trigrams
- ☐ D. 21 bigrams and 20 trigrams
- ☐ E. 19 bigrams and 18 trigrams
- ☐ F. 20 bigrams and 21 trigrams

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 9 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌘ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 10 of 20 > ⏩

Question 10

3 points ✓ Saved

Which of the following is the distance of two clusters is the distance between the two mean vectors of the clusters?

- ☐ A. Average Linkage
- ☐ B. Single Linkage
- ☒ C. Centroid Linkage
- ☐ D. Complete Linkage
- ☐ E. Euclidian Distance
- ☐ F. Hamming Distance

→ ⚠ Moving to another question will save this response.

⏪ < Question 10 of 20 > ⏩

Question 11

3 points

✓ Saved

A data scientist wants to cluster the following records based on the five attributes, X1, X2, X3, X4, and X5 directly, using the k-means algorithm with Euclidean distance as the measure without any transformation. Which of the following statement is true?

X1	X2	X3	X4	X5
4	79	68	10	862367
4	17	767	12	338490
3	41	120	13	378964
2	56	996	4	498648
3	25	815	3	913552
2	48	51	3	562023
1	49	91	7	824055
5	25	184	15	362452
3	17	603	11	573746
2	43	73	10	574701
1	87	292	8	322586
4	16	453	7	533687
4	11	429	9	152592
5	67	509	12	378566
2	76	211	3	393462
2	42	41	13	808208
2	98	657	14	630258
5	95	242	15	968338
5	25	923	11	140312
3	33	665	6	115934

- ☐ A. All the five attributes will contribute to the clustering model equally.
- ☒ B. X3 will play a dominant role in the clustering.
- ☐ C. X1 will play a dominant role in the clustering.
- ☐ D. X2 will play a dominant role in the clustering.
- ☐ E. X4 will play a dominant role in the clustering.
- ☐ F. X5 will play a dominant role in the clustering.

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

✖ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 12 of 20 ⏪ ⏩

Question 12

3 points  Saved

• In the context of text mining, what does IDF stand for, and what does it represent?

- ☐ A. IDF stands for Individual Document Factor, showing the relevance of a term within a specific document.
- ☒ B. IDF stands for Inverse Document Frequency, representing the importance of a term based on its frequency in many documents.
- ☐ C. IDF stands for In-Depth Frequency, measuring how frequently a term appears across documents.
- ☐ D. IDF stands for Information Document Factor, indicating the significance of a term within the entire document collection.
- ☐ E. IDF stands for In-Document Frequency, indicating the number of times a term occurs within a document.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 12 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌘ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 13 of 20 > ⏩

Question 13

4 points 

Which of the following conditions does NOT lead to the convergence or termination of a k-means algorithm? Please select all that apply.

- ☐ A. No centroids need to move or change.
- ☒ B. Each cluster has the similar size of items.
- ☒ C. All clusters have sufficient data points.
- ☐ D. Average within centroid distance is the same for each cluster.
- ☒ E. The number of clusters is greater than the number of data points.
- ☐ F. No data points need to change their cluster.

→ ⚠ Moving to another question will save this response.

⏪ < Question 13 of 20 > ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.
You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌵ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 14 of 20 ⏪ ⏩

Question 14

3 points  Saved

Which of the following statements best describes the Agglomerative type of Hierarchical Clustering?

- ☒ A. It starts with the points as individual clusters and merges the closest pair at each step.
- ☐ B. It starts with one, all-inclusive cluster.
- ☐ C. It creates hierarchical clusters by randomly selecting data points.
- ☐ D. It's a top-down approach, where you begin with a single, large cluster and recursively divide it into smaller clusters.
- ☐ E. It divides the dataset into clusters containing individual points.
- ☐ F. At each step, it splits a cluster until each cluster contains a point.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 14 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

✖ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 15 of 20 > ⏩

Question 15

3 points Save Answer

• Which of the following is NOT used as a column in the term-document matrix?

- ☐ A. Term occurrence
- ☐ B. TF-IDF
- ☐ C. Inverse document frequency
- ☐ D. Term frequency
- ☒ E. Bag of Words

→ ⚠ Moving to another question will save this response.

⏪ < Question 15 of 20 > ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 16 of 20 ⏪ ⏩

Question 16

3 points ✓ Saved

Which of the following statement about text mining is correct?

- ☐ A. The term-document matrix is a simple array in which each column describes a document and each row represents a particular word.
- ☒ B. N-gram considers the sequence of words in a document, but bag of words does not.
- ☐ C. N-gram means the combination of any n letters in a text document.
- ☐ D. Text mining is a directed data mining method.
- ☐ E. Tokenization is the pre-processing step for text mining that identifies sentences from a document.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 16 of 20 ⏪ ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌵ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 17 of 20 > ⏩

Question 17

3 points  Saved

• Which of the following rules has the largest confidence value?

- ☒ A. 108 out of 123 bread purchases also include milk.
- ☐ B. Among 98 transactions with hot dogs, 76 also include sweet relish.
- ☐ C. Hot dogs are included in 145 transactions and 89 of them also include apple juice.
- ☐ D. Among those transactions with sweet relish, 65.2% of them also include hot dogs.
- ☐ E. Customers who purchase hamburger buns have a 60.8% chance to purchase hot dogs.

→ ⚠ Moving to another question will save this response.

⏪ < Question 17 of 20 > ⏩

Take Test: Exam 2 Part 1

Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.
You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ < Question 18 of 20 > ⏩

Question 18

5 points  Saved

Which of the following is NOT true of a strong rule? Choose all that apply.

- ☒ A. A strong rule ($A \rightarrow B$) has the same confidence as its reversed rule ($B \rightarrow A$) and both of them meet the minimum confidence threshold.
- ☒ B. A strong rule is always an actionable rule.
- ☒ C. If a rule meets the minimum support threshold, it must be a strong rule.
- ☐ D. A strong rule must meet both the minimum support and confidence thresholds.
- ☐ E. If a rule has a quite high confidence such as 0.8 or 0.9, it must be a strong rule.
- ☐ F. A strong rule ($A \rightarrow B$) has the same support as its reversed rule ($B \rightarrow A$) and both of them meet the minimum support threshold.

→ ⚠ Moving to another question will save this response.

⏪ < Question 18 of 20 > ⏩

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

⌵ Question Completion Status:

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 19 of 20 ⏪ ⏩

Question 19

8 points ✔ Saved

Suppose that we have two strings: $A=1011001011$ and $B=1001101110$. Please compute the following distance measures. Attention: Round to the second decimal place. If a particular distance measure is not applicable, please type NA.

The Jaccard distance between A and B is

The normalized Hamming distance between A and B is

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 19 of 20 ⏪ ⏩

Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

Save and Submit

Click **Submit** to complete this assessment.

Question 20 of 20

Question 20

3 points 

Which one of the following is an indicator of a good clustering result?

- ☐ A. Each cluster has an (almost) equal number of objects.
- ☒ B. The distances between any two clusters are large while the distance between any two data points in the same cluster is small.
- ☐ C. The average within centroid distance is the smallest.
- ☐ D. The number of clusters is more than 3 but smaller than 12.
- ☐ E. The distances between any two clusters are small while the distance between any two data points in the same cluster is large.

Click **Submit** to complete this assessment.

Question 20 of 20

Save and Submit

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

Question Completion Status:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

⚠ Moving to another question will save this response.

Question 1 of 20 >

Question 1

5 points ✓ Saved

A telecommunication company develops a prediction model to predict whether or not a customer is going to churn and plans to launch a customer care program to maintain those customers who are predicted to churn. In order to test the model's performance, the following confusion matrix is generated.

		Predicted Class	
		Churn=Yes	Churn=No
Actual Class	Churn=Yes	250	50
	Churn=No	100	600

Suppose this company has calculated the cost for each prediction as below.

		Predicted Class	
		Churn=Yes	Churn=No
Actual Class	Churn=Yes	\$4	\$67
	Churn=No	\$31	\$-7

Please compute the total cost of this prediction model (just type the number without dollar signs as your answer such as 12345).

3250

⚠ Moving to another question will save this response.

Question 1 of 20 >

Question 11

3 points 

A data scientist wants to cluster the following records based on the five attributes, X1, X2, X3, X4, and X5 directly, using the k-means algorithm with Euclidean distance as the measure without any transformation. Which of the following statement is true?

X1	X2	X3	X4	X5
4	79	68	10	862367
4	17	767	12	338490
3	41	120	13	378964
2	56	996	4	498648
3	25	815	3	913552
2	48	51	3	562023
1	49	91	7	824055
5	25	184	15	362452
3	17	603	11	573746
2	43	73	10	574701
1	87	292	8	322586
4	16	453	7	533687
4	11	429	9	152592
5	67	509	12	378566
2	76	211	3	393462
2	42	41	13	808208
2	98	657	14	630258
5	95	242	15	968338
5	25	923	11	140312
3	33	665	6	115934

- ☐ A. All the five attributes will contribute to the clustering model equally.
- ☐ B. X3 will play a dominant role in the clustering.
- ☐ C. X1 will play a dominant role in the clustering.
- ☐ D. X2 will play a dominant role in the clustering.
- ☐ E. X4 will play a dominant role in the clustering.
- ☒ F. X5 will play a dominant role in the clustering.

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.
You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

★ Question Completion Status:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

⏪ ⚠ Moving to another question will save this response.

⏪ < Question 19 of 20 > ⏩

Question 19

8 points

Save Answer

Suppose that we have two strings: $A=1011001011$ and $B=1001101110$. Please compute the following distance measures. Attention: Round to the second decimal place. If a particular distance measure is not applicable, please type NA.

The Jaccard distance between A and B is

The normalized Hamming distance between A and B is

⏪ ⚠ Moving to another question will save this response.

⏪ < Question 19 of 20 > ⏩

Take Test: Exam 2 Part 1

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

★ Question Completion Status:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

⚠ Moving to another question will save this response.

⏪ ⏩ Question 19 of 20 ⏪ ⏩

Question 19

8 points ✓ Saved

Suppose that we have two strings: $A=1011001011$ and $B=1001101110$. Please compute the following distance measures. Attention: Round to the second decimal place. If a particular distance measure is not applicable, please type NA.

The Jaccard distance between A and B is

The normalized Hamming distance between A and B is

⚠ Moving to another question will save this response.

⏪ ⏩ Question 19 of 20 ⏪ ⏩

★ Test Information

Description Part 1 of Exam 2 includes 20 questions (75 points in total), including multiple choice, multiple answer, and computation questions.

You have one attempt with 80 minutes before the due time. Your grade will NOT be available during the exam period.

Instructions

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

★ Question Completion Status:



→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 18 of 20 ⏪ ⏩

Question 18

5 points  Saved

Which of the following is NOT true of a strong rule? Choose all that apply.

- ☐ A. A strong rule ($A \rightarrow B$) has the same confidence as its reversed rule ($B \rightarrow A$) and both of them meet the minimum confidence threshold.
- ☒ B. A strong rule is always an actionable rule.
- ☒ C. If a rule meets the minimum support threshold, it must be a strong rule.
- ☐ D. A strong rule must meet both the minimum support and confidence thresholds.
- ☐ E. If a rule has a quite high confidence such as 0.8 or 0.9, it must be a strong rule.
- ☐ F. A strong rule ($A \rightarrow B$) has the same support as its reversed rule ($B \rightarrow A$) and both of them meet the minimum support threshold.

→ ⚠ Moving to another question will save this response.

⏪ ⏩ Question 18 of 20 ⏪ ⏩