

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	59 minutes	4.75 out of 5

! Correct answers are hidden.

Score for this quiz: **4.75** out of 5

Submitted Dec 20 at 20:29

This attempt took 59 minutes.

Question 1

0.25 / 0.25 pts

In positively skewed data distribution, mean will be less than the median?

True

False

Question 2

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of

error outlier**Question 3****0.25 / 0.25 pts**

The objectives of Data pre-processing are

 All of the above Improve Data quality Modify data to better fit specific data mining technique Fill the missing value**Question 4****0.25 / 0.25 pts**

Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?

 Regression Classification Clustering Sequential Pattern discovery**Question 5****0.25 / 0.25 pts**

Outlier treatment can be performed using

-
- Process of Retaining, Rectifying and Removing
-
- None
-
- Process of creating dummy variables
-
- Process of creating new variables

Question 6**0.25 / 0.25 pts**

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?

-
- True
-
- False

Question 7**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The values of quartile 2 (Q2) is given by:

-
- 40
-
- 60

20 34.5**Question 8****0.25 / 0.25 pts**

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

 Sequential Pattern discovery Classification Clustering Regression**Question 9****0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

 Both the given options None of the given options Business Constraints Business Objectives

Question 10**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features



Deriving the correlation between the features



Clustering activities



Identifying outliers

Incorrect**Question 11****0 / 0.25 pts**

The Correlation coefficient between two variables was returned as -1.011. What does this infer?



None of the above



The variables are negatively correlated



The variables are independent



The variables are positively correlated

Question 12**0.25 / 0.25 pts**

Which one of the following is not part of the data preprocessing task?

- Data Classification
- Data Normalization
- Dimension Reduction
- Feature Selection

Question 13**0.25 / 0.25 pts**

Data scrubbing can be defined as

- Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
- Analyzing data to discover rules and relationship to detect violators
- Check field overloading
- Delete redundant tuples

Question 14**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The interquartile range is given by:

40 60 20 34.5**Question 15****0.25 / 0.25 pts**

Dimensionality reduction reduces the data set size by removing

 irrelevant attributes relevant attributes derived attributes composite attributes**Question 16****0.25 / 0.25 pts**

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

 27 and 5

6 and 27 11 and 6 5 and 27**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

 1.1791 0.1791 1.194 1.21**Question 18****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25

0.50 0.75**Question 19****0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

 1 and 3

 1 and 2

 2 and 3

 Indeterminate**Question 20****0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz Score: 4.75 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	40 minutes	4.75 out of 5

! Correct answers are hidden.

Score for this quiz: **4.75** out of 5

Submitted Dec 20 at 20:12

This attempt took 40 minutes.

Question 1	0.25 / 0.25 pts
Which method is suitable for data reduction	
<input checked="" type="radio"/> All of the above	
<input type="radio"/> Regression	
<input type="radio"/> Clustering	
<input type="radio"/> Histogram	

Question 2**0.25 / 0.25 pts**

Example of Continuous Data Type

-
- Food consumed by a dog (in weight)
 - Employee IDs
 - None of the above
 - Employee ratings
-

Question 3**0.25 / 0.25 pts**

Which one of the following is not part of the data preprocessing task?

-
- Data Classification
 - Data Normalization
 - Dimension Reduction
 - Feature Selection
-

Question 4**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 5

0.25 / 0.25 pts

Which of the following is an example of the predictive task of data mining?

Predicting price of a house based on historical price values and the features

Deriving the correlation between the features

Clustering activities

None of these

Incorrect

Question 6

0 / 0.25 pts

The process that is not associated with data cleaning is

segmentation

disambiguation

- deduplication
- domain consistency

Question 7**0.25 / 0.25 pts**

Which of the following are data mining activities?

- Predicting the future price of a stock based on historical data
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 8**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The interquartile range is given by:

- 40
- 60
- 20

- 34.5

Question 9**0.25 / 0.25 pts**

Which of the following activities is not a data mining task?

- Extracting the frequencies of a sound wave
- Monitoring the heart rate of a patient for abnormalities.
- Predicting the future stock price of a company using historical records.
- Monitoring and predicting failures in a hydro power plan

Question 10**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 11**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

- 4, 4, 15, 15
- 4, 4, 4, 15
- 21, 21, 25, 25
- 26, 26, 26, 34

Question 12**0.25 / 0.25 pts**

Where the use of metadata will be useful

- To avoid errors in schema integration
- Missing values
- For inconsistency
- None of the above

Question 13**0.25 / 0.25 pts**

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

True

False

Question 14**0.25 / 0.25 pts**

The correlation between two feature variables x and y is computed as 0.86. This indicates

x, y are strongly correlated positively -as x increases, y also increases

x, y are weakly correlated -as x increases, y decreases

Cannot be determined

x, y are not correlated

Question 15**0.25 / 0.25 pts**

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

smooth by bin values

smooth by bin boundaries

- smooth by bin median
- smooth by bin means

Question 16

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

- 27 and 5
- 6 and 27
- 11 and 6
- 5 and 27

Question 17

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

- 15

- 12
- 13
- 14

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

- 1 and 3
- 1 and 2
- 2 and 3
- Indeterminate

Question 19

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791 0.1791 1.194 1.21**Question 20****0.25 / 0.25 pts**

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

 Classification followed by Prediction Clustering followed by Prediction Classification followed by Association Sequential pattern discovery**Quiz Score: 4.75 out of 5**

Question 1

0.25 / 0.25 pts

What is the Interquartile range for the below set of data points:

1,1,1,3,4,5,5,6,9,11,13,14,17,18,21



11



6



8



7

Question 2

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates



x, y are strongly correlated positively -as x increases, y also increases



x, y are weakly correlated -as x increases, y decreases



Cannot be determined



x, y are not correlated

Incorrect Question 3

0 / 0.25 pts

Which of the following activities is not a data mining task?



Extracting the frequencies of a sound wave



Monitoring the heart rate of a patient for abnormalities.



Predicting the future stock price of a company using historical records.



Monitoring and predicting failures in a hydro power plant

Question 4

0.25 / 0.25 pts

Extreme values that occur infrequently are called as _____



outliers



rare values



dimensionality of data



supremum values

Question 5

0.25 / 0.25 pts

Data scrubbing can be defined as



Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections



Analyzing data to discover rules and relationship to detect violators



Check field overloading



Delete redundant tuples

Question 6

0.25 / 0.25 pts

In the following which one is not part of data transformation?



Filling missing value



Aggregation



Normalization



Smoothing

Question 7

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of



error



outlier

Question 8

0.25 / 0.25 pts

Consider a data set where most of the values pertaining to a particular feature are missing.

Which of the below is most appropriate way to handle this problem?



Perform correlation analysis and based on this, explore whether this feature can be dropped



Filling all the missing entries with some default value, say 0



Filling all the missing entries with some default value, say -99



Find the mean amongst available values and fill the missing values with this mean

Question 9

0.25 / 0.25 pts

Which statement is not TRUE regarding a data mining task:



Regression is a descriptive data mining task



Classification is a predictive data mining task



Clustering is a descriptive data mining task



Deviation detection is a predictive data mining task

Question 10

0.25 / 0.25 pts

Identify which one is Predictive Task?



Classification



Clustering



Association Rule Discovery



Sequential Pattern Discovery

Incorrect Question 11

0 / 0.25 pts

Example of Continuous Data Type



Food consumed by a dog (in weight)



Employee IDs



None of the above



Employee ratings

Question 12

0.25 / 0.25 pts

Dimensionality reduction reduces the data set size by removing



irrelevant attributes



relevant attributes



derived attributes



composite attributes

Question 13

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step



smooth by bin values



smooth by bin boundaries



smooth by bin median



smooth by bin means

Question 14

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?



Sequential Pattern discovery



Classification



Clustering



Regression

Question 15

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of



Classification



Clustering



Association Rule Mining



Regression

Question 16

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]



1.1791



0.1791



1.194



1.21

Question 17

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---



27 and 5



6 and 27



11 and 6



5 and 27

Question 18

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects



15



12



13



14

Incorrect Question 19

0 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]



1 and 3



1 and 2



2 and 3



Indeterminate

Incorrect Question 20

0 / 0.25 pts

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz 1 Results for KOTALWAR SANGAMESH

❗ Correct answers are hidden.

Score for this quiz: **4.5** out of 5

Submitted Dec 20 at 20:51

This attempt took 18 minutes.

Question 1

0.25 / 0.25 pts

Which of the following is empirically valid approximate relation among mean, median and mode?

- mean -mode $\approx=3(\text{mean} - \text{median})$
- mean -median $\approx=3(\text{mean} - \text{mode})$
- median -mode $\approx=3(\text{median} - \text{mean})$
- mean -median $\approx=(\text{mean}-\text{mode})$

Question 2

0.25 / 0.25 pts

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering ($<$, $>$)?

- True
- False

Question 3

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

-
- Sequential Pattern discovery
 - Classification
 - Clustering
 - Regression
-

Incorrect

Question 4

0 / 0.25 pts

Attributes cannot be called as:

-
- Tuple
 - Dimensions
 - Variables
 - Features
-

Question 5

0.25 / 0.25 pts

Which of the following are data mining activities?



Predicting the future price of a stock based on historical data



Monitoring diabetes profile of patients to check whether they maintain control levels or not



Computing total sales of a company



Dividing customers according to their gender

Question 6

0.25 / 0.25 pts

Consider a data set where most of the values pertaining to a particular feature are missing.

Which of the below is most appropriate way to handle this problem?



Perform correlation analysis and based on this, explore whether this feature can be dropped



Filling all the missing entries with some default value, say 0



Filling all the missing entries with some default value, say -99



Find the mean amongst available values and fill the missing values with this mean

Incorrect

Question 7

0 / 0.25 pts

Which of the following is not a data mining functionality?

- Selection and interpretation
- Classification and regression
- Characterization and Discrimination
- Clustering and Analysis

Question 8

0.25 / 0.25 pts

Data Quality problems area

- All of the above
- Noise and outliers
- Missing values
- Duplicate data

Question 9

0.25 / 0.25 pts

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test

- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering
- Association Rule Mining
- Regression

Question 11

0.25 / 0.25 pts

Which of the following data mining task is known as Market Basket Analysis?

- Association Rule analysis
- Classification
- Regression
- Outlier analysis

Question 12

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task



Identify and group the similar documents according to the context



Find the average salary of the employee in the grade B



Query a Web search engine for information about “Amazon”



Look up phone number in phone directory

Question 13

0.25 / 0.25 pts

Correlation analysis is used for



Identifying redundant attributes



Handling different data formats



Eliminating noise



Handling missing values

Question 14

0.25 / 0.25 pts

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

True

False

Question 15

0.25 / 0.25 pts

Dimensionality reduction reduces the data set size by removing

irrelevant attributes

relevant attributes

derived attributes

composite attributes

Question 16

0.25 / 0.25 pts

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

0.33

0.25

0.50

0.75

Question 17

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

-
- Classification followed by Prediction
 - Clustering followed by Prediction
 - Classification followed by Association
 - Sequential pattern discovery
-

Quiz Score: **4.5** out of 5

Question 1

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of



Classification



Clustering



Association Rule Mining



Regression

Question 2

0.25 / 0.25 pts

Correlation analysis is used for



Identifying redundant attributes



Handling different data formats



Eliminating noise



Handling missing values

Question 3

0.25 / 0.25 pts

Larger the chi-square value between two nominal attributes means



Attributes are related



Attributes are not related



None of the these



Causal relation between attributes

Incorrect Question 4

0 / 0.25 pts

Which one of the following is not an alternative to data mining?



Computational Intelligence



Knowledge Extraction



Data Dredging



Knowledge Discovery in Database (KDD)

Question 5

0.25 / 0.25 pts

The objectives of Data pre-processing are



All of the above



Improve Data quality



Modify data to better fit specific data mining technique



Fill the missing value

Question 6

0.25 / 0.25 pts

Where the use of metadata will be useful



To avoid errors in schema integration



Missing values



For inconsistency



None of the above

Question 7

0.25 / 0.25 pts

In a dataset, Hair_color is one of the attributes and it can take the following values {Red, Green, Yellow, Black}, what kind of attribute is it?



Nominal



Ordinal



Continuous



None

Question 8

0.25 / 0.25 pts

What is an Imputation?

Select the best option



Process of handling missing values in the data set



Process of identifying Outliers



Both the given options



None of the given options

Question 9

0.25 / 0.25 pts

Example of Continuous Data Type



Food consumed by a dog (in weight)



Employee IDs



None of the above



Employee ratings

Question 10

0.25 / 0.25 pts

Identify which one is Predictive Task?



Classification



Clustering



Association Rule Discovery



Sequential Pattern Discovery

Question 11

0.25 / 0.25 pts

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?



True



False

Question 12

0.25 / 0.25 pts

Which of the following are descriptive data mining activities?



Clustering



Regression



Classification



None

Question 13

0.25 / 0.25 pts

Jersey number of cricket players is?



Nominal



Ordinal



Interval



Ratio

Question 14

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The interquartile range is given by:



40



60



20



34.5

Question 15

0.25 / 0.25 pts

Which of the following is an example of the predictive task of data mining?



Predicting price of a house based on historical price values and the features



Deriving the correlation between the features



Clustering activities



None of these

Incorrect Question 16

0 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1



Incorrect



Correct

Question 17

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects



15



12



13



14

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]



1 and 3



1 and 2



2 and 3



Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---



27 and 5



6 and 27



11 and 6



5 and 27

Question 20

0.25 / 0.25 pts

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz Score: **4.5** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:51

This attempt took 26 minutes.

Question 1

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

- Sequential Pattern discovery
- Classification
- Clustering
- Regression

Question 2**0.25 / 0.25 pts**

Considering quartiles in a normal distribution, and given that Q1 =30, Q3 =70,

The interquartile range is given by $1.5 \times 40 = 60$, Is it valid?

True

False

Question 3**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

Identify and group the similar documents according to the context

Find the average salary of the employee in the grade B

Query a Web search engine for information about “Amazon”

Look up phone number in phone directory

Question 4**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 5

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The interquartile range is given by:

40

60

20

34.5

Question 6

0.25 / 0.25 pts

The objectives of Data pre-processing are

All of the above

Improve Data quality

Modify data to better fit specific data mining technique

Fill the missing value

Question 7**0.25 / 0.25 pts**

Which of the following are data mining activities?

-
- Predicting the future price of a stock based on historical data
-
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
-
- Computing total sales of a company
-
- Dividing customers according to their gender

Question 8**0.25 / 0.25 pts**

Jersey number of cricket players is?

-
- Nominal
-
- Ordinal
-
- Interval
-
- Ratio

Question 9**0.25 / 0.25 pts**

Data Quality problems area

All of the above

Noise and outliers

Missing values

Duplicate data

Question 10

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

Will be the same

σ_X will be smaller than σ_Y .

σ_Y will be smaller than σ_X

Magnitude will be the same but the sign will be different

Question 11

0.25 / 0.25 pts

Larger the chi-square value between two nominal attributes means

Attributes are related

Attributes are not related

None of the these

- Causal relation between attributes

Question 12**0.25 / 0.25 pts**

In a dataset, Hair_color is one of the attributes and it can take the following values {Red, Green, Yellow, Black}, what kind of attribute is it?

- Nominal
- Ordinal
- Continuous
- None

Question 13**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 14**0.25 / 0.25 pts**

Outlier treatment can be performed using

-
- Process of Retaining, Rectifying and Removing
 - None
 - Process of creating dummy variables
 - Process of creating new variables
-

Question 15

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

-
- smooth by bin values
 - smooth by bin boundaries
 - smooth by bin median
 - smooth by bin means
-

Question 16

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
---	--	--	--

1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 17

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

0.33

0.25

0.50

0.75

Question 19

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 20

0.25 / 0.25 pts

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00 **Points** 5 **Questions** 20
Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day **Time Limit** 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	31 minutes	4.75 out of 5

! Correct answers are hidden.

Score for this quiz: **4.75** out of 5

Submitted Dec 20 at 20:56

This attempt took 31 minutes.

Question 1	0.25 / 0.25 pts
Which method is suitable for data reduction	
<input checked="" type="radio"/> All of the above	
<input type="radio"/> Regression	
<input type="radio"/> Clustering	
<input type="radio"/> Histogram	

Question 2	0.25 / 0.25 pts

Data Quality problems area

- All of the above
- Noise and outliers
- Missing values
- Duplicate data

Question 3

0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

- Both the given options
- None of the given options
- Business Constraints
- Business Objectives

Question 4

0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as

- mean
- mode

- frequency
- None of the above

Question 5**0.25 / 0.25 pts**

Attributes cannot be called as:

- Tuple
- Dimensions
- Variables
- Features

Question 6**0.25 / 0.25 pts**

Data scrubbing can be defined as

- Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
- Analyzing data to discover rules and relationship to detect violators
- Check field overloading
- Delete redundant tuples

Question 7**0.25 / 0.25 pts**

Identify which one is Predictive Task?

- Classification
- Clustering
- Association Rule Discovery
- Sequential Pattern Discovery

Question 8**0.25 / 0.25 pts**

Which one of the following is not a challenge or issue in the data mining process?

- Extracting information from domain experts, database and system administrators
- Handling uncertainty, noise, or incompleteness of data
- Mining knowledge in multidimensional space
- d. Mining different types and new kinds of knowledge

Incorrect**Question 9****0 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation
- Classification and regression
- Characterization and Discrimination
- Clustering and Analysis

Question 10**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 11**0.25 / 0.25 pts**

Where the use of metadata will be useful

- To avoid errors in schema integration
- Missing values
- For inconsistency
- None of the above

Question 12**0.25 / 0.25 pts**

In a dataset, Hair_color is one of the attributes and it can take the following values {Red, Green, Yellow, Black}, what kind of attribute is it?

- Nominal
- Ordinal
- Continuous
- None

Question 13**0.25 / 0.25 pts**

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

True False**Question 14****0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

 True False**Question 15****0.25 / 0.25 pts**

Larger the chi-square value between two nominal attributes means

 Attributes are related Attributes are not related None of the these Causal relation between attributes**Question 16****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 17

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 18

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

- 15
- 12
- 13
- 14

Question 19**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

- 0.33
- 0.25
- 0.50
- 0.75

Question 20**0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz Score: **4.75** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1 Results for SABHARINATH B

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:08

This attempt took 17 minutes.

Question 1

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features



Deriving the correlation between the features



Clustering activities



Identifying outliers

Question 2

0.25 / 0.25 pts

Jersey number of cricket players is?



Nominal



Ordinal

- Interval
- Ratio

Question 3

0.25 / 0.25 pts

Which of the following are data mining activities?

- Predicting the future price of a stock based on historical data
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 4

0.25 / 0.25 pts

Which one of the following is not a challenge or issue in the data mining process?

- Extracting information from domain experts, database and system administrators
- Handling uncertainty, noise, or incompleteness of data

- Mining knowledge in multidimensional space
- d. Mining different types and new kinds of knowledge

Question 5

0.25 / 0.25 pts

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 6

0.25 / 0.25 pts

The process that is not associated with data cleaning is

- segmentation
- disambiguation
- deduplication
- domain consistency

Question 7

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of

 error

 outlier**Question 8**

0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as

 mean

 mode

 frequency

 None of the above**Question 9**

0.25 / 0.25 pts

The objectives of Data pre-processing are

 All of the above

- Improve Data quality
- Modify data to better fit specific data mining technique
- Fill the missing value

Question 10

0.25 / 0.25 pts

Chi-square test is suitable for

- Nominal data
- Numerical data
- Multimedia data
- Transaction data

Question 11

0.25 / 0.25 pts

Which method is suitable for data reduction

- All of the above
- Regression
- Clustering
- Histogram

Question 12

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

- Sequential Pattern discovery
- Classification
- Clustering
- Regression

Question 13

0.25 / 0.25 pts

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs
- None of the above
- Employee ratings

Question 14

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

Classification

Clustering

Association Rule Mining

Regression

Question 15

0.25 / 0.25 pts

Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?

Regression

Classification

Clustering

Sequential Pattern discovery

Question 16

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 17

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 18

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15

12

13

14

Question 19

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

-
- Classification followed by Prediction
 - Clustering followed by Prediction
 - Classification followed by Association
 - Sequential pattern discovery
-

Quiz Score: **5** out of 5

Quiz 1 Results for

! Correct answers are hidden.

Score for this quiz: **4.75** out of 5

Submitted Dec 20 at 20:48

This attempt took 33 minutes.

Question 1

0.25 / 0.25 pts

The process that is not associated with data cleaning is

- segmentation
- disambiguation
- deduplication
- domain consistency

Question 2

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 3

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of

 error outlier**Question 4**

0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as

 mean mode frequency None of the above**Question 5**

0.25 / 0.25 pts

There are two sets $X = \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41\}$

-42, -43, -44}. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

-
- Will be the same
-
- σ_X will be smaller than σ_Y .
-
- σ_Y will be smaller than σ_X
-
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

-
- 40
-
- 60
-
- 20
-
- 34.5

Question 7

0.25 / 0.25 pts

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

True False**Question 8**

0.25 / 0.25 pts

Larger the chi-square value between two nominal attributes means

 Attributes are related Attributes are not related None of the these Causal relation between attributes**Question 9**

0.25 / 0.25 pts

Data Quality problems area

 All of the above Noise and outliers Missing values Duplicate data

Question 10

0.25 / 0.25 pts

Which of the following methods can be used for eliminating redundancy in Nominal data.

-
- Chi-square test
-
- Correlation coefficient (Pearson's coefficient)
-
- Covariance
-
- Option A and B

Question 11

0.25 / 0.25 pts

Data scrubbing can be defined as

-
- Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
-
- Analyzing data to discover rules and relationship to detect violators
-
- Check field overloading
-
- Delete redundant tuples

Incorrect

Question 12

0 / 0.25 pts

The Correlation coefficient between two variables was returned as -1.011. What does this infer?

- None of the above
- The variables are negatively correlated
- The variables are independent
- The variables are positively correlated

Question 13

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

- smooth by bin values
- smooth by bin boundaries
- smooth by bin median
- smooth by bin means

Question 14

0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 15

0.25 / 0.25 pts

What is an Imputation?

Select the best option

Process of handling missing values in the data set

Process of identifying Outliers

Both the given options

None of the given options

Question 16

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1				
1	1			
0	0	1		
0	1	0	1	

Incorrect

Correct

Question 17

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15 12 13 14**Question 19**

0.25 / 0.25 pts

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 20**

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: **4.75** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	26 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 20:41

This attempt took 26 minutes.

Question 1	0.25 / 0.25 pts
Identify which one is Predictive Task?	
<input checked="" type="radio"/> Classification	
<input type="radio"/> Clustering	
<input type="radio"/> Association Rule Discovery	
<input type="radio"/> Sequential Pattern Discovery	

Question 2**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

4, 4, 15, 15

4, 4, 4, 15

21, 21, 25, 25

26, 26, 26, 34

Question 3**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 4**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

- rare values
- dimensionality of data
- supremum values

Question 5

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 6

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

- Classification
- Clustering

- Association Rule Mining
- Regression

Question 7**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 8**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"

- Look up phone number in phone directory

Question 9**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 10**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

- Process of handling missing values in the data set
- Process of identifying Outliers
- Both the given options
- None of the given options

Question 11**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 12**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

- Business Objectives

Question 13**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error

- outlier

Question 14**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

- Selection and interpretation

- Classification and regression

- Characterization and Discrimination

- Clustering and Analysis

Question 15**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

mean mode frequency None of the above**Question 16****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 17****0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1
Due Dec 21 at 01:00 Points: 20 Available: Dec 20 at 01:00 - Dec 21 at 01:00 Time Limit: 40 minutes

Instructions

Present all of the quiz in a cohesive continuous learning and share in place of the answers.

Attempt History

User	Attempts	Date	Score
UNP	1	0 hours ago	4.75 out of 5

Submit

Question 1 0.25 / 0.25 pts

Which method/s is/are valid for data reduction?

- What's above
- Aggregation
- Sampling
- Generalization

Question 2 0.25 / 0.25 pts

Data Quality problems are:

- What's above
- Data inconsistency
- Missing values
- Outliers

Question 3 0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

- Both of the options
- None of the options
- Business Context
- Business Objectives

Question 4 0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as:

- Mean
- Median
- Mode
- Range of data

Question 5 0.25 / 0.25 pts

Attribute cannot be called as:

- Type
- Dimension
- Variable
- Attribute

Question 6 0.25 / 0.25 pts

Data wrangling can be defined as:

- For anyone involved in wrangling, just make quick check to their errors and take corrective action
- Analysis which focuses on how to make data more useful
- Data Cleaning
- Data Transformation

Question 7 0.25 / 0.25 pts

Identify which one is Predictive Task?

- Descriptive
- Predictive
- Associative Rule Discovery
- Descriptive Rule Discovery

Question 8 0.25 / 0.25 pts

What one of the following is not a data mining function?

- Extracting information from data
- Identifying interesting items in a collection of data
- Mining knowledge in a collection of data
- Mining different parts and their kinds of knowledge

Question 9 0.25 / 0.25 pts

What one of the following is not a data mining function?

- Descriptive rule generation
- Descriptive analysis
- Descriptive and Predictive
- Descriptive and Predictive

Question 10 0.25 / 0.25 pts

Correlation analysis is used for:

- Describing relevant data
- Identifying relevant new features
- Describing rules
- Identifying missing values

Question 11 0.25 / 0.25 pts

What one of the instances will be used?

- To build model or perform migration
- To validate
- To predict
- None of the above

Question 12 0.25 / 0.25 pts

In a dataset, after joining two of the left tables and I can take the following values (Red, Green, Yellow, Black) what kind of attribute is it?

- Nominal
- Ordinal
- Continuous
- Ratio

Question 13 0.25 / 0.25 pts

The process of pre-processing data to transform the raw input data into a proper format suitable for further processing?

- True
- False

Question 14 0.25 / 0.25 pts

In probability density G(x)-distribution, mean will be less than the median?

- True
- False

Question 15 0.25 / 0.25 pts

Large standard error value between two normal distribution results.

- Metrics are tested
- Metrics are tested
- None of the above
- Standard deviation

Question 16 0.25 / 0.25 pts

Given a sequence of values:
9, 4, 3, 36, 36, 15, 15, 25, 25, 46, 46, 33
what is the new maximum value for 30 if it is the target (1, 2)

- 12%
- 15%
- 18%
- 20%

Question 17 0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Dissimilarity table?

Person A	1	2	3	4
Person B	1	2	3	4
1	1	2	3	4
2	2	1	3	4
3	3	4	1	2
4	4	3	2	1

Answered

Question 18 0.25 / 0.25 pts

Given two objects represented by the binary {1, 1, 0, 1, 1, 0, 1, 0} and {0, 1, 0, 1, 0, 1, 0, 1}, calculate the Spearman distance between the two objects

- 0.5
- 0.25
- 0.75
- 0.0

Question 19 0.25 / 0.25 pts

Assuming all attributes are binary symmetric, what is the Jaccard Coefficient for 1st and 2nd attributes from the given table?

Attribute	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1

Answered

Question 20 0.25 / 0.25 pts

What one of the following is referred to as a classification task?

- Classification
- Regression
- Clustering
- Dimensionality Reduction

Selection and interpretation

Quiz 1 Results for SABHARINATH B

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:08

This attempt took 17 minutes.

Question 1

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features



Deriving the correlation between the features



Clustering activities



Identifying outliers

Question 2

0.25 / 0.25 pts

Jersey number of cricket players is?



Nominal



Ordinal

- Interval
- Ratio

Question 3

0.25 / 0.25 pts

Which of the following are data mining activities?

- Predicting the future price of a stock based on historical data
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 4

0.25 / 0.25 pts

Which one of the following is not a challenge or issue in the data mining process?

- Extracting information from domain experts, database and system administrators
- Handling uncertainty, noise, or incompleteness of data

- Mining knowledge in multidimensional space
- d. Mining different types and new kinds of knowledge

Question 5

0.25 / 0.25 pts

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 6

0.25 / 0.25 pts

The process that is not associated with data cleaning is

- segmentation
- disambiguation
- deduplication
- domain consistency

Question 7

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of

 error

 outlier**Question 8**

0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as

 mean

 mode

 frequency

 None of the above**Question 9**

0.25 / 0.25 pts

The objectives of Data pre-processing are

 All of the above

- Improve Data quality
- Modify data to better fit specific data mining technique
- Fill the missing value

Question 10

0.25 / 0.25 pts

Chi-square test is suitable for

- Nominal data
- Numerical data
- Multimedia data
- Transaction data

Question 11

0.25 / 0.25 pts

Which method is suitable for data reduction

- All of the above
- Regression
- Clustering
- Histogram

Question 12

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

- Sequential Pattern discovery
- Classification
- Clustering
- Regression

Question 13

0.25 / 0.25 pts

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs
- None of the above
- Employee ratings

Question 14

0.25 / 0.25 pts

Google tries to differentiate emails as spam and non-spam, this is an example of

Classification

Clustering

Association Rule Mining

Regression

Question 15

0.25 / 0.25 pts

Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?

Regression

Classification

Clustering

Sequential Pattern discovery

Question 16

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 17

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 18

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15

12

13

14

Question 19

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

-
- Classification followed by Prediction
 - Clustering followed by Prediction
 - Classification followed by Association
 - Sequential pattern discovery
-

Quiz Score: **5** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	27 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:29

This attempt took 27 minutes.

Question 1

0.25 / 0.25 pts

Which of the following are data mining activities?

- Predicting the future price of a stock based on historical data
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 2**0.25 / 0.25 pts**

Attributes cannot be called as:

- Tuple
- Dimensions
- Variables
- Features

Question 3**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

- mean
- mode
- frequency
- None of the above

Question 4**0.25 / 0.25 pts**

Which of the following is empirically valid approximate relation among mean, median and mode?

mean -mode $\approx 3(\text{mean} - \text{median})$

mean -median $\approx 3(\text{mean} - \text{mode})$

median -mode $\approx 3(\text{median} - \text{mean})$

mean -median $\approx (\text{mean} - \text{mode})$

Question 5

0.25 / 0.25 pts

Which of the following activities is not a data mining task?

Extracting the frequencies of a sound wave

Monitoring the heart rate of a patient for abnormalities.

Predicting the future stock price of a company using historical records.

Monitoring and predicting failures in a hydro power plant

Question 6

0.25 / 0.25 pts

Which of the following data mining task is known as Market Basket Analysis?

Association Rule analysis

Classification

Regression

Outlier analysis

Question 7

0.25 / 0.25 pts

The objectives of Data pre-processing are

All of the above

Improve Data quality

Modify data to better fit specific data mining technique

Fill the missing value

Question 8

0.25 / 0.25 pts

Which of the following is an example of the predictive task of data mining?

Predicting price of a house based on historical price values and the features

Deriving the correlation between the features

- Clustering activities
- None of these

Question 9**0.25 / 0.25 pts**

Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?

- Regression
- Classification
- Clustering
- Sequential Pattern discovery

Question 10**0.25 / 0.25 pts**

Which statement is not TRUE regarding a data mining task:

- Regression is a descriptive data mining task
- Classification is a predictive data mining task
- Clustering is a descriptive data mining task
- Deviation detection is a predictive data mining task

Question 11**0.25 / 0.25 pts**

Considering quartiles in a normal distribution, and given that Q1 =30, Q3 =70,

The interquartile range is given by $1.5 \times 40 = 60$, Is it valid?

True

False

Question 12**0.25 / 0.25 pts**

Which of the following is not a Data discretization Method:

Data compression

Histogram analysis

Binning

None of the given option

Question 13**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

Selection and interpretation

- Classification and regression
- Characterization and Discrimination
- Clustering and Analysis

Question 14**0.25 / 0.25 pts**

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 15**0.25 / 0.25 pts**

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

- True
- False

Question 16**0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 17**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

0.33

0.25

0.50

0.75

Question 18**0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

 Incorrect

 Correct**Question 19****0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

 1 and 3

 1 and 2

- 2 and 3
- Indeterminate

Question 20**0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?

- Identification of areas of similar land use in an earth observation database
- Motorists who are at high risk of a car accident in the next 12 months
- Houses that are likely to rise in value in 12 months' time
- Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1 Results for KANUKANTI ANIRUDH

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:27

This attempt took 28 minutes.

Question 1	0.25 / 0.25 pts
<p>Correlation analysis is used for</p> <hr/> <p><input checked="" type="radio"/> Identifying redundant attributes</p> <hr/> <p><input type="radio"/> Handling different data formats</p> <hr/> <p><input type="radio"/> Eliminating noise</p> <hr/> <p><input type="radio"/> Handling missing values</p>	

Question 2	0.25 / 0.25 pts
<p>Which one of the following is not a challenge or issue in the data mining process?</p> <hr/> <p><input checked="" type="radio"/> Extracting information from domain experts, database and system administrators</p> <hr/> <p><input type="radio"/> Handling uncertainty, noise, or incompleteness of data</p> <hr/> <p><input type="radio"/> Mining knowledge in multidimensional space</p> <hr/> <p><input type="radio"/> d. Mining different types and new kinds of knowledge</p>	

Question 3	0.25 / 0.25 pts
<p>The process that is not associated with data cleaning is</p> <hr/> <p><input checked="" type="radio"/> segmentation</p> <hr/> <p><input type="radio"/> disambiguation</p>	

- deduplication
- domain consistency

Question 4

0.25 / 0.25 pts

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

- True
- False

Question 5

0.25 / 0.25 pts

Attributes cannot be called as:

- Tuple
- Dimensions
- Variables
- Features

Question 6

0.25 / 0.25 pts

Larger the chi-square value between two nominal attributes means

- Attributes are related
- Attributes are not related
- None of the these
- Causal relation between attributes

Question 7

0.25 / 0.25 pts

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

 Ordinal Nominal Continuous None**Question 8**

0.25 / 0.25 pts

Which of the following data mining task is known as Market Basket Analysis?

 Association Rule analysis Classification Regression Outlier analysis**Question 9**

0.25 / 0.25 pts

What is an Imputation?

Select the best option

 Process of handling missing values in the data set

- Process of identifying Outliers
- Both the given options
- None of the given options

Question 10

0.25 / 0.25 pts

Data Quality problems area

- All of the above
- Noise and outliers
- Missing values
- Duplicate data

Question 11

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

- smooth by bin values
- smooth by bin boundaries
- smooth by bin median
- smooth by bin means

Question 12

0.25 / 0.25 pts

In a dataset, Hair_color is one of the attributes and it can take the following values {Red, Green, Yellow, Black}, what kind of attribute is it?

- Nominal

Ordinal

Continuous

None

Question 13

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The values of quartile 2 (Q2) is given by:

40

60

20

34.5

Question 14

0.25 / 0.25 pts

The objectives of Data pre-processing are

All of the above

Improve Data quality

Modify data to better fit specific data mining technique

Fill the missing value

Question 15

0.25 / 0.25 pts

Which of the following are data mining activities?

Predicting the future price of a stock based on historical data

- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 16

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

- 15
- 12
- 13
- 14

Question 17

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

- 1.1791
- 0.1791
- 1.194
- 1.21

Question 18

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
Frequency	2	1	2	3	1	1	1

27 and 5

6 and 27

11 and 6

5 and 27

Question 19

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 20

0.25 / 0.25 pts

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database

Motorists who are at high risk of a car accident in the next 12 months

Houses that are likely to rise in value in 12 months' time

- Customers who are likely to buy a particular product

Quiz Score: **5** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	22 minutes	4 out of 5

! Correct answers are hidden.

Score for this quiz: 4 out of 5

Submitted Dec 20 at 21:24

This attempt took 22 minutes.

Incorrect

Question 1

0 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about “Amazon”
- Look up phone number in phone directory

Question 2**0.25 / 0.25 pts**

Attributes cannot be called as:

- Tuple
- Dimensions
- Variables
- Features

Question 3**0.25 / 0.25 pts**

Which of the following is not a Data discretization Method:

- Data compression
- Histogram analysis
- Binning
- None of the given option

Question 4**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

40 60 20 34.5**Question 5****0.25 / 0.25 pts**

Correlation analysis is used for

 Identifying redundant attributes Handling different data formats Eliminating noise Handling missing values

Incorrect

Question 6**0 / 0.25 pts**

Which of the following are data mining activities?

 Predicting the future price of a stock based on historical data
Monitoring diabetes profile of patients to check whether they maintain control levels or not

- Computing total sales of a company
- Dividing customers according to their gender

Question 7**0.25 / 0.25 pts**

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

- Sequential Pattern discovery
- Classification
- Clustering
- Regression

Question 8**0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 9**0.25 / 0.25 pts**

Consider a data set where most of the values pertaining to a particular feature are missing.

Which of the below is most appropriate way to handle this problem?



Perform correlation analysis and based on this, explore whether this feature can be dropped



Filling all the missing entries with some default value, say 0



Filling all the missing entries with some default value, say -99



Find the mean amongst available values and fill the missing values with this mean

Question 10**0.25 / 0.25 pts**

Identify which one is Predictive Task?



Classification



Clustering



Association Rule Discovery



Sequential Pattern Discovery

Question 11**0.25 / 0.25 pts**

Jersey number of cricket players is?

 Nominal Ordinal Interval Ratio

Incorrect

Question 12**0 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

 4, 4, 15, 15 4, 4, 4, 15 21, 21, 25, 25 26, 26, 26, 34

Incorrect

Question 13**0 / 0.25 pts**

Which of the following activities is not a data mining task?

- Extracting the frequencies of a sound wave
- Monitoring the heart rate of a patient for abnormalities.
- Predicting the future stock price of a company using historical records.
- Monitoring and predicting failures in a hydro power plant

Question 14**0.25 / 0.25 pts**

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 15**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features

- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 16**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

- 15
- 12
- 13
- 14

Question 17**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

- 0.33

0.25 0.50 0.75**Question 18****0.25 / 0.25 pts**

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

 27 and 5 6 and 27 11 and 6 5 and 27**Question 19****0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 20**0.25 / 0.25 pts**

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 4 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	32 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:31

This attempt took 32 minutes.

Question 1

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates



x, y are strongly correlated positively -as x increases, y also increases



x, y are weakly correlated -as x increases, y decreases



Cannot be determined



x, y are not correlated

Question 2**0.25 / 0.25 pts**

Chi-square test is suitable for

- Nominal data
- Numerical data
- Multimedia data
- Transaction data

Question 3**0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 4**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context

- Find the average salary of the employee in the grade B
- Query a Web search engine for information about “Amazon”
- Look up phone number in phone directory

Question 5**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

- 4, 4, 15, 15
- 4, 4, 4, 15
- 21, 21, 25, 25
- 26, 26, 26, 34

Question 6**0.25 / 0.25 pts**

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs
- None of the above
- Employee ratings

Question 7**0.25 / 0.25 pts**

Which of the following data mining task is known as Market Basket Analysis?

-
- Association Rule analysis
-
- Classification
-
- Regression
-
- Outlier analysis

Question 8**0.25 / 0.25 pts**

How do you understand the Problem Statement before you start your data mining activity?

-
- Both the given options
-
- None of the given options
-
- Business Constraints
-
- Business Objectives

Question 9**0.25 / 0.25 pts**

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

1- Very Unsatisfied

2- Somewhat Unsatisfied

3- Neutral

4- Somewhat Satisfied

5- Very Satisfied

What type of attribute is satisfaction here?

Ordinal

Nominal

Continuous

None

Question 10

0.25 / 0.25 pts

The salary of the employee is -20000. This is the problem of

error

outlier

Question 11

0.25 / 0.25 pts

Which of the following is not a Data discretization Method:

Data compression

- Histogram analysis
- Binning
- None of the given option

Question 12**0.25 / 0.25 pts**

Larger the chi-square value between two nominal attributes means

- Attributes are related
- Attributes are not related
- None of the these
- Causal relation between attributes

Question 13**0.25 / 0.25 pts**

Which method is suitable for data reduction

- All of the above
- Regression
- Clustering
- Histogram

Question 14**0.25 / 0.25 pts**

The process that is not associated with data cleaning is

- segmentation
- disambiguation
- deduplication
- domain consistency

Question 15**0.25 / 0.25 pts**

The objectives of Data pre-processing are

- All of the above
- Improve Data quality
- Modify data to better fit specific data mining technique
- Fill the missing value

Question 16**0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

 Incorrect Correct**Question 17****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75**Question 18****0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is

the right amount of loan. From Data Scientist's perspective, bank is performing

-
- Classification followed by Prediction
 - Clustering followed by Prediction
 - Classification followed by Association
 - Sequential pattern discovery
-

Quiz Score: **5** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	22 minutes	3.75 out of 5

! Correct answers are hidden.

Score for this quiz: 3.75 out of 5

Submitted Dec 20 at 21:12

This attempt took 22 minutes.

Incorrect

Question 1

0 / 0.25 pts

Which of the following activities is not a data mining task?

- Extracting the frequencies of a sound wave
- Monitoring the heart rate of a patient for abnormalities.
- Predicting the future stock price of a company using historical records.
- Monitoring and predicting failures in a hydro power plant

Question 2**0.25 / 0.25 pts**

Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?

-
- Regression
 - Classification
 - Clustering
 - Sequential Pattern discovery
-

Question 3**0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

-
- Computational Intelligence
 - Knowledge Extraction
 - Data Dredging
 - Knowledge Discovery in Database (KDD)
-

Incorrect**Question 4****0 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

40 60 20 34.5**Question 5****0.25 / 0.25 pts**

Which one of the following is not part of the data preprocessing task?

 Data Classification Data Normalization Dimension Reduction Feature Selection**Question 6****0.25 / 0.25 pts**

In the following which one is not part of data transformation?

 Filling missing value Aggregation

- Normalization
- Smoothing

Incorrect

Question 7

0 / 0.25 pts

Which of the following are data mining activities?

- Predicting the future price of a stock based on historical data
- Monitoring diabetes profile of patients to check whether they maintain control levels or not
- Computing total sales of a company
- Dividing customers according to their gender

Question 8

0.25 / 0.25 pts

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance

- Option A and B

Question 9**0.25 / 0.25 pts**

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

-
- True
-

- False

Question 10**0.25 / 0.25 pts**

Considering quartiles in a normal distribution, and given that $Q1 = 30$, $Q3 = 70$,

The interquartile range is given by $1.5 * 40 = 60$, Is it valid?

-
- True
-

- False

Question 11**0.25 / 0.25 pts**

There are two sets $X = \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41\}$

-42, -43, -44}. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

-
- Will be the same
-
- σ_X will be smaller than σ_Y .
-
- σ_Y will be smaller than σ_X
-
- Magnitude will be the same but the sign will be different

Question 12**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The interquartile range is given by:

-
- 40
-
- 60
-
- 20
-
- 34.5

Question 13**0.25 / 0.25 pts**

Which of the following is not a Data discretization Method:

-
- Data compression

- Histogram analysis
- Binning
- None of the given option

Question 14**0.25 / 0.25 pts**

Which of the following data mining task is known as Market Basket Analysis?

- Association Rule analysis
- Classification
- Regression
- Outlier analysis

Incorrect**Question 15****0 / 0.25 pts**

The Correlation coefficient between two variables was returned as -1.011. What does this infer?

- None of the above
- The variables are negatively correlated
- The variables are independent

- The variables are positively correlated

Question 16**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

- 0.33
- 0.25
- 0.50
- 0.75

Question 17**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

- 15
- 12
- 13

14

Incorrect

Question 18

0 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

 1 and 3

 1 and 2

 2 and 3

 Indeterminate**Question 19**

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			

1	1			
0	0	1		
0	1	0	1	

Incorrect

Correct

Question 20

0.25 / 0.25 pts

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database

Motorists who are at high risk of a car accident in the next 12 months

Houses that are likely to rise in value in 12 months' time

Customers who are likely to buy a particular product

Quiz Score: **3.75** out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	58 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:35

This attempt took 58 minutes.

Question 1

0.25 / 0.25 pts

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?

True

False

Question 2

0.25 / 0.25 pts

What is the Interquartile range for the below set of data points:

1,1,1,3,4,5,5,6,9,11,13,14,17,18,21

11

6

8

7

Question 3

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70

The values of quartile 2 (Q2) is given by:

40

60

20

34.5

Question 4

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"
- Look up phone number in phone directory

Question 5**0.25 / 0.25 pts**

Which of the following activities is not a data mining task?

- Extracting the frequencies of a sound wave
- Monitoring the heart rate of a patient for abnormalities.
- Predicting the future stock price of a company using historical records.
- Monitoring and predicting failures in a hydro power plant

Question 6**0.25 / 0.25 pts**

Larger the chi-square value between two nominal attributes means

- Attributes are related
- Attributes are not related
- None of the these

- Causal relation between attributes

Question 7**0.25 / 0.25 pts**

Outlier treatment can be performed using

- Process of Retaining, Rectifying and Removing
- None
- Process of creating dummy variables
- Process of creating new variables

Question 8**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

- error
- outlier

Question 9**0.25 / 0.25 pts**

The Correlation coefficient between two variables was returned as -1.011. What does this infer?

- None of the above
- The variables are negatively correlated
- The variables are independent
- The variables are positively correlated

Question 10**0.25 / 0.25 pts**

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs
- None of the above
- Employee ratings

Question 11**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features

- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 12**0.25 / 0.25 pts**

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 13**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

- mean
- mode
- frequency

- None of the above

Question 14**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

-
- 4, 4, 15, 15
-
- 4, 4, 4, 15
-
- 21, 21, 25, 25
-
- 26, 26, 26, 34

Question 15**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

-
- Selection and interpretation
-
- Classification and regression
-
- Characterization and Discrimination
-
- Clustering and Analysis

Question 16**0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 17**0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 18**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

 15 12 13 14**Question 19****0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

 0.33 0.25 0.50 0.75

Question 20**0.25 / 0.25 pts**

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

- Classification followed by Prediction
- Clustering followed by Prediction
- Classification followed by Association
- Sequential pattern discovery

Quiz Score: 5 out of 5

Quiz 1 Results for PRADYUT ROY

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:36

This attempt took 31 minutes.

Question 1

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

- smooth by bin values
- smooth by bin boundaries
- smooth by bin median
- smooth by bin means

Question 2

0.25 / 0.25 pts

Which of the following activities is not a data mining task?

- Extracting the frequencies of a sound wave
- Monitoring the heart rate of a patient for abnormalities.
- Predicting the future stock price of a company using historical records.
- Monitoring and predicting failures in a hydro power plant

Question 3

0.25 / 0.25 pts

There are two sets $X=\{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

-
- Will be the same
 - σ_X will be smaller than σ_Y .
 - σ_Y will be smaller than σ_X
 - Magnitude will be the same but the sign will be different

Question 4

0.25 / 0.25 pts

In the following which one is not part of data transformation?

-
- Filling missing value
 - Aggregation
 - Normalization
 - Smoothing

Question 5

0.25 / 0.25 pts

Which of the following data mining task is known as Market Basket Analysis?

-
- Association Rule analysis
 - Classification
 - Regression
 - Outlier analysis
-

Question 6

0.25 / 0.25 pts

Data Quality problems area

-
- All of the above
 - Noise and outliers
 - Missing values
 - Duplicate data
-

Question 7

0.25 / 0.25 pts

Consider a data set where most of the values pertaining to a particular feature are missing.

Which of the below is most appropriate way to handle this problem?



Perform correlation analysis and based on this, explore whether this feature can be dropped



Filling all the missing entries with some default value, say 0



Filling all the missing entries with some default value, say -99



Find the mean amongst available values and fill the missing values with this mean

Question 8

0.25 / 0.25 pts

Which of the following are descriptive data mining activities?



Clustering



Regression



Classification



None

Question 9

0.25 / 0.25 pts

The sum of observed data points divided by the number of data records is called as

- mean
- mode
- frequency
- None of the above

Question 10

0.25 / 0.25 pts

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 11

0.25 / 0.25 pts

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs

- None of the above
- Employee ratings

Question 12

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 13

0.25 / 0.25 pts

Considering quartiles in a normal distribution, and given that Q1 =30, Q3 =70,

The interquartile range is given by $1.5*40 =60$, Is it valid?

- True
- False

Question 14

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

 40

 60

 20

 34.5**Question 15**

0.25 / 0.25 pts

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?

 True

 False**Question 16**

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum

distance between the two objects

15

12

13

14

Question 17

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 18

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 19

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 20**0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database



Motorists who are at high risk of a car accident in the next 12 months



Houses that are likely to rise in value in 12 months' time



Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	11 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:42

This attempt took 11 minutes.

Question 1

0.25 / 0.25 pts

There are two sets $X = \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24\}$ and $Y = \{-30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44\}$. What is TRUE about the standard deviations of X and Y i.e. σ_X and σ_Y respectively?

- Will be the same
- σ_X will be smaller than σ_Y .
- σ_Y will be smaller than σ_X
- Magnitude will be the same but the sign will be different

Question 2**0.25 / 0.25 pts**

Which of the following is empirically valid approximate relation among mean, median and mode?

- mean -mode $\approx=3(\text{mean} - \text{median})$
- mean -median $\approx=3(\text{mean} - \text{mode})$
- median -mode $\approx=3(\text{median} - \text{mean})$
- mean -median $\approx=(\text{mean}-\text{mode})$

Question 3**0.25 / 0.25 pts**

Sorted data (attribute values) for price are 4, 8, 9, 15, 21, 21, 21, 24, 25, 26, 28, 29, 34. Identify which is NOT a bin smoothed by boundaries?

- 4, 4, 15, 15
- 4, 4, 4, 15
- 21, 21, 25, 25
- 26, 26, 26, 34

Question 4**0.25 / 0.25 pts**

Which of the following is not a Data discretization Method:

- Data compression
- Histogram analysis
- Binning
- None of the given option

Question 5**0.25 / 0.25 pts**

Data scrubbing can be defined as

- Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
- Analyzing data to discover rules and relationship to detect violators
- Check field overloading
- Delete redundant tuples

Question 6**0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

- Chi-square test
- Correlation coefficient (Pearson's coefficient)
- Covariance
- Option A and B

Question 7

0.25 / 0.25 pts

In point-of-sale transaction sequences "(Shoes), (Racket, Racketball) followed by (Sports_Jacket)" pattern is discovered. Which data mining task could detect this pattern?

- Sequential Pattern discovery
- Classification
- Clustering
- Regression

Question 8

0.25 / 0.25 pts

Larger the chi-square value between two nominal attributes means

- Attributes are related
- Attributes are not related

- None of the these
- Causal relation between attributes

Question 9**0.25 / 0.25 pts**

The process that is not associated with data cleaning is

- segmentation
- disambiguation
- deduplication
- domain consistency

Question 10**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 11**0.25 / 0.25 pts**

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?

-
- True
-
- False

Question 12**0.25 / 0.25 pts**

Example of Continuous Data Type

-
- Food consumed by a dog (in weight)
-
- Employee IDs
-
- None of the above
-
- Employee ratings

Question 13**0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging
- Knowledge Discovery in Database (KDD)

Question 14

0.25 / 0.25 pts

In positively skewed data distribution, mean will be less than the median?

- True
- False

Question 15

0.25 / 0.25 pts

Which of the following is "NOT" an example of descriptive tasks of data mining?

- Predicting price of a house based on historical price values and the features
- Deriving the correlation between the features
- Clustering activities

- Identifying outliers

Question 16**0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

-
- 1 and 3
-

- 1 and 2
-

- 2 and 3
-

- Indeterminate
-

Question 17**0.25 / 0.25 pts**

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

-
- 1.1791
-

- 0.1791
-

1.194 1.21**Question 18**

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

 27 and 5 6 and 27 11 and 6 5 and 27**Question 19**

0.25 / 0.25 pts

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

 15

12 13 14**Question 20****0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database

 Motorists who are at high risk of a car accident in the next 12 months Houses that are likely to rise in value in 12 months' time Customers who are likely to buy a particular product**Quiz Score: 5 out of 5**

Quiz 1

Due Dec 21 at 8:30am

Points 5

Questions 20

Available Dec 20 at 8:30am - Dec 21 at 8:30am 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	8 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: **5** out of 5

Submitted Dec 20 at 11:13am

This attempt took 8 minutes.

Question 1

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 2**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

 40 60 20 34.5**Question 3****0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

 mean mode frequency None of the above**Question 4****0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

-
- Computational Intelligence
 - Knowledge Extraction
 - Data Dredging
 - Knowledge Discovery in Database (KDD)
-

Question 5**0.25 / 0.25 pts**

Which of the following are data mining activities?

-
- Predicting the future price of a stock based on historical data
 - Monitoring diabetes profile of patients to check whether they maintain control levels or not
 - Computing total sales of a company
 - Dividing customers according to their gender
-

Question 6**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

rare values

dimensionality of data

supremum values

Question 7

0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 8

0.25 / 0.25 pts

Example of Continuous Data Type

Food consumed by a dog (in weight)

Employee IDs

None of the above

- Employee ratings

Question 9**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features

Deriving the correlation between the features

Clustering activities

Identifying outliers

Question 10**0.25 / 0.25 pts**

The process that is not associated with data cleaning is

segmentation

disambiguation

deduplication

domain consistency

Question 11**0.25 / 0.25 pts**

Outlier treatment can be performed using

-
- Process of Retaining, Rectifying and Removing
 - None
 - Process of creating dummy variables
 - Process of creating new variables
-

Question 12**0.25 / 0.25 pts**

Chi-square test is suitable for

-
- Nominal data
 - Numerical data
 - Multimedia data
 - Transaction data
-

Question 13**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 14

0.25 / 0.25 pts

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

- Ordinal
- Nominal
- Continuous
- None

Question 15

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

-
- Identify and group the similar documents according to the context
-
- Find the average salary of the employee in the grade B
-
- Query a Web search engine for information about "Amazon"
-
- Look up phone number in phone directory

Question 16**0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

-
- Incorrect
-

- Correct

Question 17**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

0.33

0.25

0.50

0.75

Question 18

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 19**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15

12

13

14

Question 20**0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database

Motorists who are at high risk of a car accident in the next 12 months

Houses that are likely to rise in value in 12 months' time

Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	31 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: **5 out of 5**

Submitted Dec 20 at 21:41

This attempt took 31 minutes.

Question 1

0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 2**0.25 / 0.25 pts**

The salary of the employee is -20000. This is the problem of

error

outlier

Question 3**0.25 / 0.25 pts**

Dimensionality reduction reduces the data set size by removing

irrelevant attributes

relevant attributes

derived attributes

composite attributes

Question 4**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

Identify and group the similar documents according to the context

Find the average salary of the employee in the grade B

- Query a Web search engine for information about “Amazon”
- Look up phone number in phone directory

Question 5**0.25 / 0.25 pts**

The objectives of Data pre-processing are

- All of the above
- Improve Data quality
- Modify data to better fit specific data mining technique
- Fill the missing value

Question 6**0.25 / 0.25 pts**

The nominal attributes are just labeled with valid operations as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?

- True
- False

Question 7**0.25 / 0.25 pts**

Which of the following is not a data mining functionality?

-
- Selection and interpretation
 - Classification and regression
 - Characterization and Discrimination
 - Clustering and Analysis
-

Question 8**0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

-
- mean
 - mode
 - frequency
 - None of the above
-

Question 9**0.25 / 0.25 pts**

Larger the chi-square value between two nominal attributes means

- Attributes are related
- Attributes are not related
- None of the these
- Causal relation between attributes

Question 10

0.25 / 0.25 pts

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

- 40
- 60
- 20
- 34.5

Question 11

0.25 / 0.25 pts

Which of the following is an example of the predictive task of data mining?

-
- Predicting price of a house based on historical price values and the features

- Deriving the correlation between the features
- Clustering activities
- None of these

Question 12**0.25 / 0.25 pts**

Which of the following is not a Data discretization Method:

- Data compression
- Histogram analysis
- Binning
- None of the given option

Question 13**0.25 / 0.25 pts**

Which method is suitable for data reduction

- All of the above
- Regression
- Clustering
- Histogram

Question 14**0.25 / 0.25 pts**

Data scrubbing can be defined as



- Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
-



- Analyzing data to discover rules and relationship to detect violators
-



- Check field overloading
-



- Delete redundant tuples

Question 15**0.25 / 0.25 pts**

Example of Continuous Data Type



- Food consumed by a dog (in weight)
-



- Employee IDs
-



- None of the above
-



- Employee ratings

Question 16**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15

12

13

14

Question 17

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 18**0.25 / 0.25 pts**

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19**0.25 / 0.25 pts**

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 20

0.25 / 0.25 pts

Which of the following is unlikely to be a classification task?



Identification of areas of similar land use in an earth observation database

Motorists who are at high risk of a car accident in the next 12 months

Houses that are likely to rise in value in 12 months' time

Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 8:30am

Points 5

Questions 20

Available Dec 20 at 8:30am - Dec 21 at 8:30am 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	8 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: **5** out of 5

Submitted Dec 20 at 11:13am

This attempt took 8 minutes.

Question 1

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 2**0.25 / 0.25 pts**

Consider the sorted list of data values given by: 10,20, 30, 40, 50, 60, 70
The values of quartile 2 (Q2) is given by:

 40 60 20 34.5**Question 3****0.25 / 0.25 pts**

The sum of observed data points divided by the number of data records is called as

 mean mode frequency None of the above**Question 4****0.25 / 0.25 pts**

Which one of the following is not an alternative to data mining?

-
- Computational Intelligence
 - Knowledge Extraction
 - Data Dredging
 - Knowledge Discovery in Database (KDD)
-

Question 5**0.25 / 0.25 pts**

Which of the following are data mining activities?

-
- Predicting the future price of a stock based on historical data
 - Monitoring diabetes profile of patients to check whether they maintain control levels or not
 - Computing total sales of a company
 - Dividing customers according to their gender
-

Question 6**0.25 / 0.25 pts**

Extreme values that occur infrequently are called as _____

outliers

rare values

dimensionality of data

supremum values

Question 7

0.25 / 0.25 pts

How do you understand the Problem Statement before you start your data mining activity?

Both the given options

None of the given options

Business Constraints

Business Objectives

Question 8

0.25 / 0.25 pts

Example of Continuous Data Type

Food consumed by a dog (in weight)

Employee IDs

None of the above

- Employee ratings

Question 9**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features

Deriving the correlation between the features

Clustering activities

Identifying outliers

Question 10**0.25 / 0.25 pts**

The process that is not associated with data cleaning is

segmentation

disambiguation

deduplication

domain consistency

Question 11**0.25 / 0.25 pts**

Outlier treatment can be performed using

- Process of Retaining, Rectifying and Removing
- None
- Process of creating dummy variables
- Process of creating new variables

Question 12**0.25 / 0.25 pts**

Chi-square test is suitable for

- Nominal data
- Numerical data
- Multimedia data
- Transaction data

Question 13**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 14

0.25 / 0.25 pts

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

- Ordinal
- Nominal
- Continuous
- None

Question 15

0.25 / 0.25 pts

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about "Amazon"
- Look up phone number in phone directory

Question 16**0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

- Incorrect
- Correct

Question 17**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

0.33

0.25

0.50

0.75

Question 18

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 19**0.25 / 0.25 pts**

Given two objects represented by the tuples (21, 12, 3, 17, 48, 11, 82, 41, 35) and (34, 5, 13, 3, 57, 26, 69, 55, 27), calculate the Supremum distance between the two objects

15

12

13

14

Question 20**0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database

Motorists who are at high risk of a car accident in the next 12 months

Houses that are likely to rise in value in 12 months' time

Customers who are likely to buy a particular product

Quiz Score: 5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	15 minutes	4.5 out of 5

! Correct answers are hidden.

Score for this quiz: **4.5** out of 5

Submitted Dec 20 at 21:24

This attempt took 15 minutes.

Question 1

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 2**0.25 / 0.25 pts**

The process of pre-processing is to transform the raw input data into a proper format suitable for further processing?

 True

 False**Question 3****0.25 / 0.25 pts**

Which of the following methods can be used for eliminating redundancy in Nominal data.

 Chi-square test

 Correlation coefficient (Pearson's coefficient)

 Covariance

 Option A and B**Incorrect****Question 4****0 / 0.25 pts**

Chi-square test is suitable for

 Nominal data

Numerical data

Multimedia data

Transaction data

Question 5

0.25 / 0.25 pts

Dimensionality reduction reduces the data set size by removing

irrelevant attributes

relevant attributes

derived attributes

composite attributes

Question 6

0.25 / 0.25 pts

Consider a data set where most of the values pertaining to a particular feature are missing.

Which of the below is most appropriate way to handle this problem?

Perform correlation analysis and based on this, explore whether this feature can be dropped

Filling all the missing entries with some default value, say 0

- Filling all the missing entries with some default value, say -99
- Find the mean amongst available values and fill the missing values with this mean

Question 7**0.25 / 0.25 pts**

Considering quartiles in a normal distribution, and given that Q1 =30, Q3 =70,

The interquartile range is given by $1.5 \times 40 = 60$, Is it valid?

- True
- False

Question 8**0.25 / 0.25 pts**

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

- smooth by bin values
- smooth by bin boundaries
- smooth by bin median
- smooth by bin means

Question 9**0.25 / 0.25 pts**

In positively skewed data distribution, mean will be less than the median?

True

False

Question 10**0.25 / 0.25 pts**

Which one of the following is not a challenge or issue in the data mining process?

Extracting information from domain experts, database and system administrators

Handling uncertainty, noise, or incompleteness of data

Mining knowledge in multidimensional space

d. Mining different types and new kinds of knowledge

Question 11**0.25 / 0.25 pts**

Which of the following is "NOT" an example of descriptive tasks of data mining?



Predicting price of a house based on historical price values and the features

- Deriving the correlation between the features
- Clustering activities
- Identifying outliers

Question 12

0.25 / 0.25 pts

Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where

- 1- Very Unsatisfied
- 2- Somewhat Unsatisfied
- 3- Neutral
- 4- Somewhat Satisfied
- 5- Very Satisfied

What type of attribute is satisfaction here?

- Ordinal

- Nominal

- Continuous

- None

Question 13**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 14**0.25 / 0.25 pts**

Identify which one is Predictive Task?

- Classification
- Clustering
- Association Rule Discovery
- Sequential Pattern Discovery

Question 15**0.25 / 0.25 pts**

Data scrubbing can be defined as



Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections

- Analyzing data to discover rules and relationship to detect violators
- Check field overloading
- Delete redundant tuples

Incorrect

Question 16

0 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

- 1 and 3

- 1 and 2

- 2 and 3

- Indeterminate

Question 17

0.25 / 0.25 pts

Given a sequence of values

3, 4, 4, 7, 10, 12, 15, 21, 25, 33, 45, 48, 70

what is the min-max normalized value for 15 in the range [1, 2]

1.1791

0.1791

1.194

1.21

Question 18

0.25 / 0.25 pts

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
----------------	---	---	---	----	----	----	----

Frequency	2	1	2	3	1	1	1
-----------	---	---	---	---	---	---	---

27 and 5

6 and 27

11 and 6

5 and 27

Question 19**0.25 / 0.25 pts**

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

 Incorrect

 Correct**Question 20****0.25 / 0.25 pts**

Which of the following is unlikely to be a classification task?

Identification of areas of similar land use in an earth observation database

 Motorists who are at high risk of a car accident in the next 12 months

 Houses that are likely to rise in value in 12 months' time

 Customers who are likely to buy a particular product

Quiz Score: 4.5 out of 5

Quiz 1

Due Dec 21 at 19:00

Points 5

Questions 20

Available Dec 20 at 19:00 - Dec 21 at 19:00 1 day

Time Limit 60 Minutes

Instructions

Purpose of the quiz is to validate continuous learning and observe grasp of the concepts.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	38 minutes	5 out of 5

! Correct answers are hidden.

Score for this quiz: 5 out of 5

Submitted Dec 20 at 21:48

This attempt took 38 minutes.

Question 1	0.25 / 0.25 pts
Which of the following activities is not a data mining task?	
<input checked="" type="radio"/> Extracting the frequencies of a sound wave	
<input type="radio"/> Monitoring the heart rate of a patient for abnormalities.	
<input type="radio"/> Predicting the future stock price of a company using historical records.	
<input type="radio"/> Monitoring and predicting failures in a hydro power plan	

Question 2**0.25 / 0.25 pts**

Correlation analysis is used for

- Identifying redundant attributes
- Handling different data formats
- Eliminating noise
- Handling missing values

Question 3**0.25 / 0.25 pts**

Which one of the following statement is reflecting data mining task

- Identify and group the similar documents according to the context
- Find the average salary of the employee in the grade B
- Query a Web search engine for information about “Amazon”
- Look up phone number in phone directory

Question 4**0.25 / 0.25 pts**

Chi-square test is suitable for

- Nominal data
- Numerical data
- Multimedia data
- Transaction data

Question 5

0.25 / 0.25 pts

The objectives of Data pre-processing are

- All of the above
- Improve Data quality
- Modify data to better fit specific data mining technique
- Fill the missing value

Question 6

0.25 / 0.25 pts

Which one of the following is not an alternative to data mining?

- Computational Intelligence
- Knowledge Extraction
- Data Dredging

- Knowledge Discovery in Database (KDD)

Question 7**0.25 / 0.25 pts**

Which of the following are descriptive data mining activities?

- Clustering
- Regression
- Classification
- None

Question 8**0.25 / 0.25 pts**

Example of Continuous Data Type

- Food consumed by a dog (in weight)
- Employee IDs
- None of the above
- Employee ratings

Question 9**0.25 / 0.25 pts**

What is an Imputation?

Select the best option

-
- Process of handling missing values in the data set
-
- Process of identifying Outliers
-
- Both the given options
-
- None of the given options

Question 10**0.25 / 0.25 pts**

Identify which one is Predictive Task?

-
- Classification
-
- Clustering
-
- Association Rule Discovery
-
- Sequential Pattern Discovery

Question 11**0.25 / 0.25 pts**

Which of the following is an example of the predictive task of data mining?



Predicting price of a house based on historical price values and the features

- Deriving the correlation between the features
- Clustering activities
- None of these

Question 12

0.25 / 0.25 pts

The correlation between two feature variables x and y is computed as 0.86. This indicates

- x, y are strongly correlated positively -as x increases, y also increases
- x, y are weakly correlated -as x increases, y decreases
- Cannot be determined
- x, y are not correlated

Question 13

0.25 / 0.25 pts

Which one of the following is not a challenge or issue in the data mining process?



Extracting information from domain experts, database and system administrators

- Handling uncertainty, noise, or incompleteness of data
- Mining knowledge in multidimensional space
- d. Mining different types and new kinds of knowledge

Question 14

0.25 / 0.25 pts

In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step

- smooth by bin values
- smooth by bin boundaries
- smooth by bin median
- smooth by bin means

Question 15

0.25 / 0.25 pts

Extreme values that occur infrequently are called as _____

- outliers

- rare values
- dimensionality of data
- supremum values

Question 16**0.25 / 0.25 pts**

Assuming all attributes are binary asymmetric; what is the Jaccard Coefficient for IT and FIN departments from the given table?

Department	Attribute1	Attribute2	Attribute3	Attribute4
IT	Y	Y	N	N
FIN	N	Y	Y	N

- 0.33
- 0.25
- 0.50
- 0.75

Question 17**0.25 / 0.25 pts**

What's the range and IQR for the below data ?

Goals per game	3	6	7	10	11	13	30
Frequency	2	1	2	3	1	1	1

27 and 5

6 and 27

11 and 6

5 and 27

Question 18

0.25 / 0.25 pts

Given the following vectors, find the pair with maximum cosine similarity.

Vector 1 = [2, 7, 1, 4]

Vector 2 = [3, 8, 1, 4]

Vector 3 = [4, 14, 2, 8]

1 and 3

1 and 2

2 and 3

Indeterminate

Question 19

0.25 / 0.25 pts

For the given records in the table, is similarity matrix correct for the Gender attribute?

Person Id	1	2	3	4
Gender	M	M	F	M

1			
1	1		
0	0	1	
0	1	0	1

Incorrect

Correct

Question 20

0.25 / 0.25 pts

A customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist's perspective, bank is performing

Classification followed by Prediction

Clustering followed by Prediction

Classification followed by Association

Sequential pattern discovery

Quiz Score: 5 out of 5

		IMPORTANT: PLEASE MARK THE ANSWER BOLD, AS THE ORDER MIGHT CHANGE, AND CHECK FOR DUPLICACY						
Question 1	What is the interquartile range for the below set of data points 1,1,1,3,4,5,5,6,9,11,13,14,17,18,21	Q1=3, Q2=6, Q3=14 Q3-Q1=11	Question 11	The process of pre-processing is to transform the raw input data into a proper format suitable for further processing	A			
	A 11 B 6 C 8 D 7		A TRUE B FALSE C D					
Question 2	Example of Continous Data Type		Question 12	Swiggy wants customers to provide their satisfaction feedback in a scale of 1-5 where 1-Very Unsatisfied 2-Somewhat Satisfied 3-Neutral 4-Somewhat Satisfied 5-Very Satisfied What type of attribute is satisfaction here?	Ordinal			
	A Food consumed by a dog(in weight) B Employee IDs C None of the above D Employee ratings	A	A Ordinal B Nominal C Continuous D None					
Question 3	Which of the following activities is not a data mining task	A	Question 13	Method suitable for data reduction-	A			
	A Extracting the frequencies of a sound wave B Monitoring the heart rate of a patient for abnormalities C Predicting the future stock price of a company using historical records D Monitoring and predicting failures in hydro power plan		A All B Regression C Clustering D Histogram					
Question 4	Extreme values that occur infrequently are called as ---	A	Question 14	In the following which one is not part of data transformation	A			
	A Outliers B Rare Values C Dimensionality of data D SuperumValues		A Filling missing value B Aggregation C Normalization D Smoothing					
Question 5	Sorted data(attribute-values) for price 4,8,9,15,21,21,24,25,26,28,29,34. Identify which is NOT a bin smoothed by boundaries		Question 15	Identify which one is predictive task				
	A 4,4,15,15 B 4,4,4,15 C 21,21,25,25 D 26,26,26,34		A Classification B Clustering C Association Rule Discovery D Sequential Pattern discovery					
Question 6	Which of the following can be used for eliminating redundancy in Nominal Data	A	Question 16	Assuming all attributes are binary asymmetric, what is Jaccard coefficient for IT and FIN departments from the given table	A		0.3333333333	
	A Chi-square test B Correlation coefficient(Pearson's coefficient) C Covariance D Both A and B		A 0.33 B 0.25 C 0.5 D 0.75	Department	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Question 7	Google tries to differentiate emails as spam and non-spam, this is an example of	A	Question 17	Given two objects represented by the tuples (21,12,3,17,48,11,82,41,35) and (34,5,13,3,57,26,69,55,27), calculate the supremum distance between the two objects	A 15 B 12 C 13 D 14			
Question 8	Outlier treatment can be performed using	A	Question 18	Where the use of metadata will be useful	A To avoid errors in schema integration B Missing values C For inconsistency D None of the above			
	A Process of Retaining, Rectifying and Removing B None C Process of creating dummy variables D Process of creating new variables							
Question 10	Considering quartiles in a normal distribution and given that Q1=30, Q3=70. The interquartile range is given by 1.5*40. Is it valid?	A	Question 19	Given the sequence of values 3,4,7,10,12,15,21,25,33,45,48,70. what is the min-max normalized value for 15 in the range [1,2]	A 1.1791 B 0.1791 C 1.194 D 1.21	1.179104478		
Question 1	The salary of employee is -20000. This is the problem of	Error (A)	Question 20	Chi-square test is suitable for	Nominal			
	A error B outlier C D			A Nominal data B Numerical data C Multimedia data D Transaction data				
Question 9	The correlation between two feature variable x and y is computed as 0.86. This indicates	A	Question 21	How do you understand the problem statement before you start your data mining activity?				
	A x, y are strongly correlated positively, as x increases y also increases B C D			A Both the given options B None of the given options C Business constraints D Business objectives				
Question 2	Which one of the following statement is reflecting data mining task	A	Question 22	Given the following vectors, find the pair with maximum cosine similarity Vector 1 = [2,7,1,4] Vector 2 = [3,8,1,4] Vector 3 = [4,14,2,8]	A 1 and 3 B 1 and 2 C 2 and 3 D Indeterminate	I V1 =8.367 I V2 =9.487 I V3 =16.733 sim(V1,V2)=0.9826 sim(V1,V3)=0.999 sim(V2,V3)=0.9929		
Question 24	Which of the following is "NOT" an example of descriptive tasks of data mining	A	Question 23	Consider the sorted list of data values given by: 10,20,30,40,50,60,70 The values of quartile 2(Q2) is given by	A 40 B 60 C 20 D 34.5			
Question 26	A Customer walks into a modern bank for obtaining loan. The bank wants to assess if loan can be given to the customer and if so, what is the right amount of loan. From Data Scientist perspective, bank is performing	A	Question 25	Which of the following is unlikely to be a classification task?				
	A Classification followed by prediction B Clustering followed by prediction			A Identification of areas of similar land use in an earth observation database B Motorists who are at high risk of a car accident in the next 12 months				

	C	Classification followed by association		C	Houses that are likely to rise in value in 12 months time			
	D	Sequential pattern discovery		D	Customers who are likely to buy a particular product			
					Given two objects represented by tuples (21,12,3,17,48,11,82,41,35) and (34,5,13,3,57,26,69,55,27), calculate the Supremum distance between the two objects			
Question 28		Data scrubbing can be defined as Use simple domain knowledge(eg postal cod, spell-check) to detect errors and make corrections	Question 27				15	
	A	Analysing data to discover rules and relationships to detect violators		A	15			
	B	Check field overloading		B	12			
	C	Delete redundant tuples		C	13			
	D			D	14			
Question 29		The process that is not associated with data cleaning is segmentation	Question 30		which of the following are descriptive data mining activities			
	A	disambiguation		A	Clustering			
	B	deduplication		B	Regression			
	C	domain consistency		C	Classification			
	D			D	None			
Question 31		Which of the following is NOT an example of descriptive tasks of data mining? Predicting price of a house based on historical price values and the features	Question 32		In point-of-sale transaction sequences "(Shoes),(Racket,Racketball)" followed by "(Sports,Jacket)" pattern is discovered. Which data mining task could detect this pattern?			
	A	Deriving the correlation between the features		A	Sequential Pattern Discovery			
	B	Clustering Activities		B	Classification			
	C	Identifying Outliers		C	Clustering			
	D			D	Regression			
Question 33		Correlation analysis is used for Identifying redundant attribute	Question 34		What's the range and IQR for the below data Goals per game 3 6 7 10 11 13 30 Frequency: 2 1 2 3 1. 1. 1			
	A	Handling different data formats		A	27 and 5			
	B	Eliminating noise		B	6 and 27			
	C	Handling missing values		C	11 and 6			
	D			D	5 and 27			
Question 35		There are two sets X=[10,11,12,13,14,15,16,17,18,19,20,21,22,23,24] and Y=[-30,-31,-32,-33,-34,-35,-36,-37,-38,-39,-40,-41,-42,-43,-44]. What is TRUE about the standard deviations of X and Y, i.e ox and oy respectively ox=4.47213 oy=4.47213	Question 36					
	A	Will be the same		A	mean - mode == 3(mean - median)			
	B	ox will be smaller than oy		B	mean - median == 3(mean - mode)			
	C	oy will be smaller than ox		C	median - mode == 3(median - mode)			
	D	Magnitude will be the same, but sign will be different		D	mean - median ~=(mean - mode)			
Question 37		In point-of-sale transaction sequences "(Shoes),(Racket,Racketball)", followed by "(Sports,Jacket)" pattern is discovered. Which data mining task could detect this pattern? Sequential Pattern discovery	Question 38		In Binning, we first sort data and partition into (equal-frequency) bins, then which of the following is not a valid step			
	A	Classification		A	smooth by bin values			
	B	Clustering		B	smooth by bin boundaries			
	C	Regression		C	smooth by bin median			
	D			D	smooth by bin means			
Question 39		The nominal attributes are just labelled with valid operation as equals, and not equals whereas on the ordinal attributes the values provide enough information on ordering (<, >)?	Question 40		In positively skewed data distribution, mean will be less than the median?			
	A	TRUE		A	TRUE			
	B	FALSE		B	FALSE			
	C			C				
	D			D				
Question 41		Which one of the following is not an alternative to data mining? Computational intelligence	Question 42		The objectives of data pre-processing are			
	A	Knowledge Extraction		A	Improve Data quality			
	B	Data Dredging		B	Modify data to better fit specific data mining technique			
	C	Knowledge Discovery in Database (KDD)		C	Fill the missing value			
	D			D	All of the above			
Question 43		The process that is not associated with data cleaning is Segmentation	Question 44		Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?			
	A	Disambiguation		A	Regression			
	B	Deduplication		B	Classification			
	C	Domain Consistency		C	Clustering			
	D			D	Sequential Pattern Discovery			
Question 45		Which of the following is not a challenge or issue in the data mining process? Extracting information from domain experts, database, and system administrator	Question 46		The correlation coefficient between two variables was returned as -1.011. What does this infer?			
	A	Handling uncertainty, noise or incompleteness of data		A	Variables are negatively correlated			
	B	Missing knowledge in multidimensional space		B	Variables are independent			
	C	Mining different types and new kind of knowledge		C	Variables are positively correlated			
	D			D	None of the above			
Question 47		The sum of observed data points divided by the number of data records is called as Mean	Question 48		Which of the following is not a data mining functionality?			
	A	Mode		A	Selection and interpretation			
	B	Frequency		B	Classification and regression			
	C	None of the above		C	Characterization and Discrimination			
	D			D	Clustering and Analysis			
Question 49		Which of the following statement is reflecting data mining task? Identify and group the similar documents according to the context	Question 50		Consider a data set where most of the values pertaining to a particular feature are missing. Which of the below is most appropriate way to handle this problem?			
	A	Find the average salary of the employee in the grade B		A	Perform correlation analysis and based on this, explore whether this feature can be dropped			
	B	Query a web search engine for information about Amazon		B	Filling all the missing entries with some default value, say 0			
	C	Lookup phone number in phone directory		C	Filling all the missing entries with some default value, say -99			
	D			D	Find the mean amongst available values and fill the missing values with this mean			
Question 51		In a dataset, Hair_color is one of the attributes and it can take the following values (Red, Green, Yellow, Black), what kind of attribute is it? Nominal	Question 52		Data Quality Problems are			
	A	Ordinal		A	Noise and outliers			
	B	Continuous		B	Missing Values			
	C	None		C	Duplicate Data			
	D			D	All of the above			
Question 53		What is an Imputation? Process of handling missing values in the data set	Question 54		Attributes cannot be called as			
	A			A	Tuple			

B		B	Dimensions
C		C	Variables
D		D	Features
Question 55	Which one of the following is not part of the data processing task?	Question 56	Larger the chi-square value between two nominal attributes means
A	Data Classification	A	Attributes are related
B	Data Normalization	B	Attributes are not related
C	Dimension Reduction	C	Causal relation between attributes
D	Feature Selection	D	None of the above
Question 57	Dimentionality reduction reduces data set by removing	Question 57	Jersey number of cricket players is?
A	irrelevant attributes	A	Nominal
B		B	Ordinal
C		C	Interval
D		D	Ratio
Question 58		Question 59	Which of the following is not a Data discretization?
A		A	Data Compression
B		B	Histogram Analysis
C		C	Binning
D		D	None of the given option
Question 60		Question 61	
A		A	
B		B	
C		C	
D		D	
Question			
A			
B			
C			
D			