Multiple Choice Questions:

Q1. Which of the following is not part of the KDD process?

A Data cleaning
B Data Mining
C Data Integration
D Data Encryption
E Data Transformation

Q2 Which of the following is a measurement of data quality?

A Accuracy
B Completeness
C Timeliness
D Reliability
E All of the above

Q3 Which of the following is not a data type?

A Nominal
B Binary
C Discrete
D Random
E Ordinal

Q4 In KDD and data mining, noisy data is referred to as _____.

A repeated data.
B complex data.
C meta data.
D random errors in database.

Q5 _____refers to the process of deriving high-quality information from text.

A. Text Mining.
B. Image Mining.
C. Database Mining.
D. Multimedia Mining.

Q6 Consider discretizing a continuous attribute whose values are listed below:

3, 4, 5, 10, 12, 21, 32, 43, 44, 46, 52, 59, 63

Using equal-width partitioning and five bins, how many values are there in the first bin?

(63-3)/5 = 60/5 = 12

Bin 1: 3-15 :3,4,5,10,12

A. 1
B. 2
C. 3
D. 4
E. 5

Q7 Which of the following are known as qualitative data types?

(i) nominal
(ii) interval
(iii) ordinal
(iv) discrete
(v) continuous

A. (i) and (ii).
B. (ii) and (iii)
C. (i) and (iii).
D. (iii) and (iv)
E (iv) and (v)

Q8 Consider discretizing a continuous attribute whose values are listed below:

3, 4, 5, 10, 20, 32, 43, 44, 46, 52, 59, 61

Which of the following number of bins is not possible for using equidepth bins?

A. 2
B. 3
C. 4
D. 5
E. 6

Q9 Suppose a group of 12 students with the test scores listed as follows:

19, 71, 48, 63, 35, 85, 69, 81, 72, 88, 100, 95

By partitioning them into three bins using equiwidth method and smoothing by bins boundaries, how many data items will be in the second bin?

19,35,48,63,69,71,72,81,85,88,95,100

A 1
B 2
C 3

D 4
E 5
(100-19)/3 = 81/3 = 27

19 – 46

47-74

Q10 This step of the KDD process model deals with noisy data.

A Data Integration
B Data pre-processing
C Data transformation
D Data mining
E Data Interpretation

Q11 Which one of the following is not part of pre-processing:

A Data cleaning
B Data integration
C Data transformation
D Data reduction
E Data visualisation

Descriptive /Problem Solving Questions

Q1) Give an example of each data type below:

Nominal

Ordinal

Binary

Discrete

Continuous

Q2) Use 5 bins to partition the following group of data based on Equal-Depth.

33, 35, 16, 21, 36, 73, 45, 22, 22, 20, 25, 25, 30, 13, 33, 16, 19, 25, 20, 35, 15, 35, 46, 52, 40.

Smooth the data within each bin using bin boundaries.

Answers:

Q1. Which of the following is not part of the KDD process.

A Data cleaning
B Data Mining
C Data Integration
D Data Encryption
E Data Transformation

Q2 Which of the following is a measurement of data quality?

A Accuracy
B Completeness
C Timeliness
D Reliability
E All of the above

Q3 Which of the following is not a data type?

A Nominal
B Binary
C Discrete
D Random
E Ordinal

Q4 In KDD and data mining, noisy data is referred to as _____.

A repeated data.
B complex data.
C meta data.
D random errors in database.

Q5 _____refers to the process of deriving high-quality information from text.

A. Text Mining.
B. Image Mining.
C. Database Mining.

D. Multimedia Mining.

Q6 Consider discretizing a continuous attribute whose values are listed below:

3, 4, 5, 10, 12, 21, 32, 43, 44, 46, 52, 59, 63

W = 63-3/5 60/5 = 12

Bin 1 : 3-15

Bin 2 : 16-28

Bin 3: 29-41

Bin 4: 42-54

Bin 5: 55- 67

Using equal-width partitioning and five bins, how many values are there in the first bin?

A. 1
B. 2
C. 3
D. 4
E. 5

Q7 Which of the following are known as qualitative data types?

(i) nominal
(ii) interval
(iii) ordinal
(iv) discrete
(v) continuous

A. (i) and (ii).
B. (ii) and (iii)
C. (i) and (iii).
D. (iii) and (iv)
E (iv) and (v)

Q8 Consider discretizing a continuous attribute whose values are listed below:

3, 4, 5, 10, 20, 32, 43, 44, 46, 52, 59, 61

Which of the following number of bins is not possible for using equidepth bins?

A. 2
B. 3
C. 4
D. 5

E. 6

Q9 Suppose a group of 12 students with the test scores listed as follows:

19, 71, 48, 63, 35, 85, 69, 81, 72, 88, 100, 95

By partitioning them into three bins using equiwidth method and smoothing by bins boundaries, how many data items will be in the second bin?

19,35,48,63,69,71,72,81,85,88,95,100 100-19/3 81/3 27

19-46

47-74

75-102

A 1
B 2
C 3
D 4
E 5

Q10 This step of the KDD process model deals with noisy data.

A Data Integration
B Data pre-processing
C Data transformation
D Data mining
E Data Interpretation

Q11 Which of the following is not part of pre-processing:

A Data cleaning
B Data integration
C Data transformation
D Data reduction
E Data visualisation

Part B: Descriptive /Problem Solving

Q1) Give an example of each data type below:

Nominal

Ordinal

Binary

Discrete

Continuous

Nominal: Hair Colour -> {black, brown, grey, red |
Ordinal size: {small, medium, large}
Binary {good, bad|
Discrete {28, 30, 31}
Continuous { 1.24, 32.4, 4.34,5.67 }

Q2) Use 5 bins to partition the following group of data based on Equal-Depth.

33, 35, 16, 21, 36, 73, 45, 22, 22, 20, 25, 25, 30, 13, 33, 16, 19, 25, 20, 35, 15, 35, 46, 52, 40.

Smooth the data within each bin using bin boundaries.

Equal Depth (5 bins)
13, 15, 16, 16, 19
20, 20, 21, 22, 22
25, 25, 25, 30, 33
33, 35, 35, 35, 36
40, 45, 46, 52, 73
Smooth by Bin Boundaries
13, 13, 13, 13, 19
20, 20, 20, 22, 22
25, 25, 25, 33, 33
33, 36, 36, 36, 36
40, 40, 40, 40, 70