|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **SCHOOL OF INFORMATION SCIENCE**

**MAKE UP EXAMINATION – JAN 2023**

**Date: 20-JAN-2023**

**Time: 1:00PM to 4:00PM**

**Max Marks: 100**

**Weightage:50%**

**Course Code:** BCA 211

**Course Name:** DATA MINING

**Program:** BCA

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*

**Part A [Memory Recall Questions]**

**Answer all the Questions. Each question carries 2 marks. (15Qx 2M= 30M)**

1. a. Explain Decision Trees. (C.O.4) [Knowledge]

b. If dangerous fires are rare (1%) but smoke is fairly common (10%) due to barbecues, and 90% of dangerous fires make smoke then what is the Probability of dangerous Fire when there is Smoke. (C.O.4) [Knowledge]

 c. Explain how Rapid Miner tool is used in Data Mining. (C.O.5) [Knowledge]

 d. In Context of Data Mining, describe what is NOISE? (C.O.3) [Knowledge]

e. Write any 4 operations performed in Rapid Miner tool. (C.O.5) [Knowledge]

f. Height and width are example for which type of data? (C.O.2) [Knowledge]

A. Finite B. Discrete C. Continuous D. None of the above

g. If the data about seismic activity in japan is given and the magnitude of the next earthquake has to be predicted. Which of the following is suitable? (C.O.1) [Knowledge]

A. Classification B. Unsupervised Learning

C. Supervised Learning D. Dimensionality Reduction

h. Which of the following is NOT an example of ordinal attributes? (C.O.1) [Knowledge]

A. Ordered numbers B. Movie ratings C. Military ranks D. Zip codes

i. Find min-max normalization by setting min=0 and max=1 for the following data:

200,300, 400, 600, 1000 (C.O.1) [Knowledge]

A. Original data 200,300, 400, 600, 1000 Normalized data 0, 0.125, 0.25, 0.5, 1

B. Original data 200,300, 400, 600, 1000 Normalized data 0.125, 0, 0.25, 1, 0.5

C. Original data 200,300, 400, 600, 1000 Normalized data 1, 0.25, 0.5,0, 0.25

D. Original data 200,300, 400, 600, 1000 Normalized data 0, 0.25, 0.5,1, 0.25

j. Which of the following Data transformation normalizes data using division by 10 power k?

 A. Decimal scaling B. Min-max C. Z score D. None of these (C.O.1) [Knowledge]

k. \_\_ is an essential process where intelligent methods are applied to extract data patterns. (C.O.2) [Knowledge]

A. Data transformation B. Data mining C. Data selection D. Text mining

l. Which of the following is an example of sequence data? (C.O.1) [Knowledge]

A. weather forecast B. genomic data C. data matrix D. market basket data

m. Let μ(mean) = 54,000, σ(standard deviation) = 16,000, then Z-score normalization value for 45200 is: A. 0.55 B. 0.54 C. -0.55 D. 0.56 (C.O.3) [Knowledge]

n. Data selection can be defined as (C.O.1) [Knowledge]

A. The actual discovery phase of a knowledge discovery process

B. The stage of selecting the right data for a KDD process.

C. A subject-oriented integrated time variant non-volatile collection of data in support of anagement.

D. None of these.

o. Given X=(2,3) and Y=(3,4), Euclidiean distance between the two points

A. 1.414 B. 1.44 C. 1.325 D. 1.432 (C.O.3) [Knowledge]

**Part B [Thought Provoking Questions]**

**Answer all the Questions. Each question carries 10 marks. (4Qx10M=40M)**

2. Along with the advantages and disadvantages, discuss in detail k-Means clustering with an example. (CO4) [Comprehension]

3. a. Use the Min-Max normalization to normalize, by setting max=10, min=5 the following

 group of data: 70, 55, 75, 35. Calculate the normalized data for all the given numbers.

(CO2) [Comprehension]

 b. You are planning a picnic today, but the morning is cloudy

* 70% of all rainy days start off cloudy!
* But cloudy mornings are common (about 60% of days start cloudy)
* And this is usually a dry month (only 6 of 30 days tend to be rainy, or 20%)

Predict the chance of rain during the day using Baye’s theorem. (CO4) [Comprehension]

4. A data mining process need to carried out for prediction of some value. Discuss the challenges we might dace during the process. (CO1) [Comprehension]

5. Data Transformation of data using Discretization need to be done on some data. Identify the types of discretization techniques for the data and elaborate on the same. (CO2) [Comprehension]

**Part C [Problem Solving Questions]**

**Answer all the Questions. Each question carries 15 marks. (2Qx15M=30M)**

6. Given a set of samples:

S = (26, N), (0, Y), (16, N), (12, N), (18, Y), (28, N), (24,N), (4,Y), (16, N)

Partition S into two intervals S1 & S2 with 21 & 14 as the split points. Find the Information Gain and find the best split. (CO2) [Application]

7. Using Apriori algorithm, formulate and validate the association rules for the transaction data given below considering Minimum Support Count = 2 and Minimum Confidence = 60%

 (CO3) [Application]

