GIFT UNIVERSITY

G U J R A N W A L A

(Chartered by the Govt. of the Punjab, Recognized by HEC)

EXA-115

Department of Computer Science

Data Mining (DS-306A)

Mid Term Examination
Fall-2022

Instructor: Dr. Muhammad Faheem

Time: 1 Hr 20 Minutes **Total Marks: 40**

Candidate Name: _____ **Candidate Roll No:** _____

Instructions to Candidates:

- Candidates are required to sit on the seats assigned to them by the invigilators.
- Do not open this question paper until you have been told to do so by the Invigilator.
- Please fill in exam specific details in space provided (both Question Paper and Answer Sheet).
- This is a Closed Book Exam. "Closed book examinations'" refer to examinations where the candidate may not take into the examination room any study materials (including textbooks, study guides, lecture notes, printed notes from web pages, handwritten notes and any audio/visual aid).
- **There are four questions. Attempt all questions.**
- Do not write anything on question paper except Name and Roll Number.

Question 1: Explain the difference between classification, regression analysis, cluster analysis, and outlier analysis. **(7.5 Mark**

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. It predicts categorical (discrete, unordered) labels.

Regression, unlike classification, is a process to model continuous-valued functions. It is used to predict missing or unavailable numerical data values rather than (discrete) class labels.

Clustering analyzes data objects without consulting a known class label. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. Each cluster that is formed can be viewed as a class of objects. Clustering can also facilitate taxonomy formation, that is, the organization of observations into a hierarchy of classes that group similar events together.

Outlier analysis is the analysis of outliers, which are objects that do not comply with the general behavior or model of the data. Examples include fraud detection based on a large dataset of credit card transactions.

Question 2: Briefly define the difference among the data matrix and the dissimilarity matrix using examples. Explain the dissimilarity between asymmetric binary attributes. **(7.5 Marks)**

Refer to Section 2.4

Question 3: Suppose that a HEC tested the faculty experience and number of publications data for 18 randomly selected faculty members with the following results: **(10 Marks)**

a) **Calculate the mean, median, and standard deviation of experience and publications.**
b) **Draw the boxplots for experience and publication.**

Experience	3	3	5	5	6	7	8	9	9
Publications	2	8	6	10	8	12	15	12	14

Experience	10	10	12	14	14	17	18	20	22
Publications	13	15	20	18	16	24	26	30	16

Solved in Classroom.

Question 4: Find the frequent items of the transactional data for an Electronics branch using Apriori Algorithm with the support count of 2 and confidence 20%. Explain each operations step wise. The table is shown below. **(15 Marks)**

TID	List of Items_ID
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1,I2, I3

Refer Example 6.3 from book

End of Question Paper.

*** SCRATCH SHEET ***

*** SCRATCH SHEET ***