

# Maulana Azad National Urdu University

B.Tech : VI Semester Examination, June 2021

Paper : BTCS611PET : Data Mining and Data Warehousing

Time : 3 hrs

Marks : 70

ہدایات:

یہ پرچہ سوالات دو حصوں پر مشتمل ہے: حصہ اول اور حصہ دوم۔ ہر جواب کے لیے لفظوں کی تعداد اشارہ ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں 10 سوالات ہیں، اس میں سے طالب علم کو کوئی 08 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً سو (100) لفظوں پر مشتمل ہے ہر سوال کے لیے 05 نمبرات مختص ہیں۔  
(8 x 5 = 40 Marks)
2. حصہ دوم میں 05 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 03 سوال کا جواب دینا ہے۔ سوال کا جواب تقریباً ڈھائی سو (250) لفظوں پر مشتمل ہے۔ سوال کے لیے 10 نمبرات مختص ہیں۔  
(10x3 = 30 Marks)

## حصہ اول

1. Knowledge Data Discovery کے Process میں Data Mining میں استعمال ہونیوالے Steps کی وضاحت کیجیے۔
2. مندرجہ ذیل Objects کے درمیان فرق کو بیان کرو۔

Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

- (a) Compute the Euclidean distance between the two objects.
  - (b) Compute the Manhattan distance between the two objects.
  - (c) Compute the Minkowski distance between the two objects, using  $h = 3$ .
  - (d) Compute the supremum distance between the two objects.
3. OLAP اور OLTP Systems کے درمیان فرق کو بیان کریں۔
  4. Data Warehouse اور Database کے درمیان فرق کو واضح کریے۔
  5.  $I = \{a, b, c, d\}$  کے لیے ایک Item Set Lattice کا خاکہ بنائیے۔
  6. "Association Rule Discovery" کی وضاحت کریں۔
  7. Classification Model کو بنانے کی ترکیب بیان کریں۔ ساتھ ہی اس کے General Approach کی وضاحت کریں۔
  8. Classifier کی Performance کو Evaluate کرنے کے Methods لکھیں۔
  9. مندرجہ ذیل Clustering طریقوں کو مثالوں کے ساتھ مختصر طور پر بیان کریں۔  
a. hierarchical methods  
b. partitioning methods
  10. مندرجہ ذیل کے مطابق ہر ایک Clustering Algorithms کی وضاحت کریں۔

- (1) shapes of clusters that can be determined;
- (2) input parameters that must be specified; and
- (3) limitations. (a) k-means (b) k-medoids (c) CLARA

## حصہ دوم

.11 Data Preprocessing کے Steps کی وضاحت کریں۔

.12 OLAP Operations کے بارے میں تفصیل سے بیان کریں۔...

.13 ایک ڈیٹا بیس میں 5 Transaction ہیں

min sup = 60% اور min conf=80% ہیں۔

TID	Items
T100	M, O, N, K, E, Y
T200	D, O, N, K, E, Y
T300	M, A, K, E
T400	M, U, C, K, Y
T500	C, O, O, K, I, E

Apriori Algorithm کا استعمال کر کے Frequent Itemsets کی تلاش کیجیے۔

.14 Decision Tree Classification میں شامل ہونے والے تمام Steps کی وضاحت کریں۔

.15 مان لیجیے کہ: Data Mining کا کام Points (x,y) جو Location کو بتاتے ہیں تین Clusters میں بتاتا ہے۔

A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9) کے Points دیے گئے ہیں۔

اور Euclidean Distance کو Distance Function سمجھائیے۔ ابتدائی طور پر ہر Cluster کا Centre بالترتیب A1, B1

اور C1 کو Assign کیا گیا ہے؟ K-means الگورتھم لکھیں اور آخری تین کلسٹرز تلاش کریں۔

☆☆☆