

Maulana Azad National Urdu University

MCA IV Semester Examination - May - 2018

Paper - MMCA402PCT : Data Warehousing and Data Mining

Time : 3 hrs

Marks : 70

ہدایات:

یہ پرچہ سوالات تین حصوں پر مشتمل ہے: حصہ اول، حصہ دوم، حصہ سوم۔ ہر جواب کے لئے لفظوں کی تعداد اشارہ ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں 10 لازمی سوالات ہیں جو کہ معروضی سوالات/خالی جگہ پُر کرنا/مختصر جواب والے سوالات ہیں۔ ہر سوال کا جواب لازمی ہے۔ ہر سوال کے لیے 1 نمبر مختص ہے۔  
(10 x 1 = 10 Marks)

2. حصہ دوم میں آٹھ سوالات ہیں، اور اس میں طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً دو سو (200) لفظوں پر مشتمل ہے۔ ہر سوال کے لیے 6 نمبرات مختص ہیں۔  
(5 x 6 = 30 Marks)

3. حصہ سوم میں پانچ سوالات ہیں۔ اس میں سے طالب علم کو کوئی تین سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً پانچ سو (500) لفظوں پر مشتمل ہے۔ ہر سوال کے لیے 10 نمبرات مختص ہیں۔  
(3 x 10 = 30 Marks)

## حصہ اول

سوال نمبر : 1

- (i) Quality of Data کی وضاحت کریں۔
- (ii) Data Mining کے چار اہم (Tasks) کام بتائیں۔
- (iii)  $i = \{a, b, c\}$  کے لیے ایک Lattice Item Set کا خاکہ (Figure) اتاریں۔
- (iv) II-Class Problem کے لیے Confusion Matrix بنائیے۔
- (v) Support اور Confidence کی وضاحت کریں۔
- (vi) Entropy اور GINI Index کی وضاحت کریں۔
- (vii) ایک Cluster جس میں (6,2), (2,3), (1,1) 3-Dimensional Points ہیں اس کا Centroid معلوم کیجیے۔
- (viii) Dendogram کیا ہے؟
- (ix) Lattice کی وضاحت کیجیے اور Spatial Statistical Analysis کی مثالیں دیجیے۔
- (x) Data Mining اور Spatial Data Mining کے درمیان فرق واضح کیجیے۔

## حصہ دوم

- (2) KDD Process کو خاکہ (Figure) کے ذریعہ تفصیل سے سمجھائیے۔
- (3) Slicing اور Rollup Operations کو مثال کے ذریعہ سمجھائیے۔
- (4) Support Vectors کیا ہیں اور اس میں Maximal Margin کے بارے میں سمجھائیے۔
- (5) Spatial Data Mining کے Systematic Structure کا خاکہ (Figure) ڈالیے۔
- (6) Agglomerative Clustering Algorithm کے بارے میں سمجھائیے۔
- (7) Classification اور Clustering کے درمیان فرق واضح کیجیے۔
- (8) کسی بھی دو Data Reduction Techniques کے بارے میں لکھیے۔
- (9) دیے گئے Supports کے مطابق

$$S(\{e\}) = 0.8; S(\{b,d\}) = 1$$

$$S(\{b,d,e\}) = 0.8$$

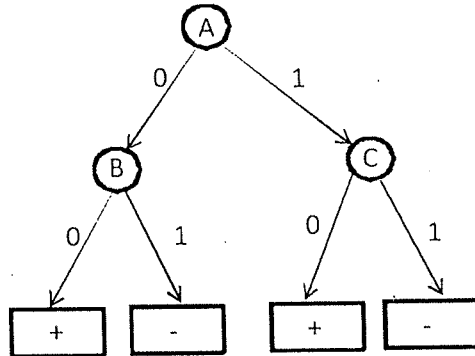
Association Rules کا Confidence معلوم کیجیے۔

$$\{b,d\} \longrightarrow \{e\};$$

$$\{e\} \longrightarrow \{b,d\}$$

## حصہ سوم

- (10) Data Pre-processing کی کیا اہمیت ہے؟ Data Transmission Strategies کی وضاحت کیجیے۔
- (11) دیے گئے 3-frequent Item Sets کے مطابق
- $\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,4\}, \{1,3,5\}, \{2,3,4\}, \{2,3,5\}, \{3,4,5\}$
- مان لیجیے کہ Data Set میں پانچ Item Sets ہی ہیں Fk-1XF1 Merging Strategy کو استعمال کرتے ہوئے Candidate Generation Procedure کے بھی چار Item Sets معلوم کیجیے۔
- (12) دیے گئے Decision-Tree کو دیکھتے ہوئے Pessimistic Approach, Optimistic Approach اور Validation Set استعمال کرتے ہوئے دیے گئے Tree کا Generalization Error Rate معلوم کیجیے۔



Training:

Instance	A	B	C	Class
1	0	0	0	+
2	0	0	1	+
3	0	1	0	+
4	0	1	1	-
5	1	0	0	+
6	1	0	0	+
7	1	1	0	-
8	1	0	1	+
9	1	1	0	-
10	1	1	0	-

Validation

Instance	A	B	C	Class
11	0	0	0	+
12	0	1	1	+
13	1	1	0	+
14	1	0	1	-
15	1	0	0	+

(13)

	X	Y
A1	2	10
A2	2	5
A3	8	4
A4	5	8
A5	7	5
A6	6	4
A7	1	2
A8	4	9

مان لیجیے کہ  $K=3$  ہے اور شروعات (Initially) میں Clusters کو مندرجہ (Following) Points دیے (Assign) گئے ہیں۔

$$C1 = \{A1, A2, A3\} \quad C2 = \{A4, A5, A6\} \quad C3 = \{A7, A8\}$$

Shortest Distance (Euclidean Distance) کو استعمال کرتے ہوئے K-means Algorithm کو لاگو (Apply) کیجیے۔

14. Spatial Data Mining کے اہم Tasks کو تفصیل سے سمجھائیے۔