

Maulana Azad National Urdu University

B.Tech : VI Semester Examination, June 2021

Paper : BTCS604PCT : Data Warehousing & Data Mining (Old)

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. Data Mining میں استعمال ہونے والے Steps کی وضاحت کیجیے۔
2. $I = \{a, b, c, d\}$ کے لیے ایک Item Set Lattice کی وضاحت کرتے ہوئے Draw کریں۔
3. Classification Model کی Performance کو Evaluate کریں۔
4. Bayes Theorem کی وضاحت کریں۔
5. Divisive Clustering اور Agglomerative Clustering کے درمیان فرق کریں۔
6. "Automatic Classification of Web Documents" کے بارے میں وضاحت کریں۔
7. Dendrogram کیا ہے یہ کس طرح تعمیر کیا گیا ہے؟
8. Data Cleaning کے دوران Missing Value بھرنے کے کتنے طریقے ہیں؟
9. Data Discrimination اور Data Summarization کے درمیان مثال کے ساتھ فرق کیجیے۔
10. Noise Data کو Smooth کیسے کریں۔

حصہ دوم

11. Pre-Processing Methods کی وضاحت کریں۔
12. Apriori Algorithm کی وضاحت کرنے کے ساتھ ہی دیے گئے ڈیٹا سٹ پر Apply کریں۔ Apply کرنے کے بعد کل کتنے Future Frequent Item Sets بن سکتے ہیں؟

TID	Items
T100	i1, i3, i4
T200	i2, i3, i5
T300	i1, i2, i3, i5
T400	i2, i5

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

.14 دیے ہوئے 8 Objects کے لیے

	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

ابتدائی طور پر $K=3$ مندرجہ ذیل clusters کو Assigned ہوئے ہیں۔

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

K-means algorithm کو Apply کرتے ہیں (Shortest Distance (Euclidean Distance) کی مدد سے Clusters بناتے ہوئے اس کے سبھی Steps لکھیں۔

.15 مندرجہ ذیل کے بارے میں مختصر لکھیے۔

(a) Multimedia Data Mining

(b) Web Usage Mining

☆☆☆