Maulana Azad National Urdu University

MCA IV Semester Examination - May - 2019

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 اشارۃً 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، اوراس میں طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً دوسو **(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) Data set کے اقسام لکھیں۔

(ii) Pattern اور Knowledge کی وضاحت کریں۔

(iii) item set I={a,b,c} درج ذیل کے لیے Lattice Draw کریں۔

(iv) ETL Tools لکھیں۔

(v) Gini Measure اور Entropy کی وضاحت کریں۔

(vi) Classification کی وضاحت کریں۔

(vii) STING کو Expand کریں۔

(viii) CLARA کیا ہے؟

(ix) Data Characterization کے Output کا Form .... ہوگا

(x) Brute-Force Method میں Rule Generation فارمولہ لکھیں۔

حصہ دوم

(2) Automatic Generation کے Hierarchy Concept کے لیے Algorithm یا Pseudo Code لکھیں جو
Nominal Database ہو اور Number of Distinct Values کے Attributes کے Schema میں ہو۔

(3) Apriori Principle کا مثال کے ساتھ وضاحت کریں۔

(4) Rule Based Classifier پر مختصر نوٹ لکھیں۔

(5) Clustering کے مختلف Partitioning Method کی وضاحت کریں۔

(6) Multimedia Data Mining پر مختصر نوٹ لکھیں۔

(7) Apriori Algorithm لکھیں۔

(8) Data Cleaning کے دوران Missing Value بھرنے کے کتنے طریقے ہیں؟

(9) Noise Data کو Smooth کیسے کریں۔

حصہ سوم

(10) Data Pre-processing Methods کی وضاحت کریں۔

(11) OLAP Operations کا Multi-dimensional ڈاٹا پر مثال کے ساتھ بحث کریں۔

(12) Binary Classification Problem کے لیے دیے گئے Table میں Training Examples کو غور کیجیے۔

Instance	a1	a2	a3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

(a Positive Class کے Respect میں اس Collection کے Training Examples کا Entropy کیا ہے۔

(b دیے گئے Training Examples سے a1 اور a2 کا Information Gain معلوم کیجیے۔

(13) مندرجہ ذیل Table کو دیکھتے ہوئے Data کو دو Clusters میں جوڑیے۔ Algorithm کو Initialize کرنے کے لیے Object 1&3 کو ایک Cluster میں اور Object 2&4 کو دوسرے Cluster میں ڈالیے۔ Algorithm کے Steps کو تفصیل سے سمجھایئے۔

Object	X1	X1
1	1	2
2	8	6
3	6	8
4	2	2

.14 مندرجہ ذیل کی وضاحت کریں۔

(a Automation Classification of Web Documents

(b Web Usage Mining

☆☆☆