مولانا آزاد نیشنل اردو یونیورسٹی

سمسٹر امتحانات **Semester Examination April/May 2026**

Programme Name: B.Tech (Computer Science) پروگرام: بی ٹیک

Semester: 6th سمسٹر:

Paper & Code: Data Mining and Data Warehousing (BTCS611PET) مضمون مع کوڈ

گھنٹے 3 Hrs وقت: :Time Maximum Marks 70 کل نمبر

ہدایات:

یہ پرچہ 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 مختصر جواب والے سوالات ہیں. ہر سوال کا جواب دینا لازمی ہے ہر سوال کے لئے 1 نمبر مختص ہے.

(Marks 10 =1 x 10)

2. حصہ دوم میں 8 سوالات ہیں، اس 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 لئے 6 نمبرات مختص ہیں

(Marks 30 =6 x 5)

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 لئے 10 نمبرات مختص ہیں.

(Marks 30 = 10 x 3)

حصہ اول

سوال (1) ہر تصور (concept) پر ایک مختصر نوٹ لکھیں۔

a۔ Data Mining کی کوئی دو مضبوط (strong) تعریفیں لکھیں۔

b۔ درج ذیل تجزیہ Data Mining task ہے یا نہیں:

i۔ "کسی کمپنی کے مستقبل کے stock price کی پیش گوئی کرنا historical records کی مدد سے"

ii۔ "کسی کمپنی کے customers کو ان کے gender کے مطابق تقسیم کرنا"

c۔ Jaccard Coefficient کی تعریف کریں۔

d- Classification Analysis کے applications لکھیں۔

e- مختلف classifiers کے نام درج کریں۔

f- Gini Index کیا ہے؟

g- Dendrogram کی تعریف کریں۔

h- IQR (Interquartile Range) کا formula لکھیں۔

i- Metric Data Objects لکھیں۔

j- درج ذیل کے لیے attribute types اور category کے نام بتائیں:

i- Height

ii- Gender

حصہ دوم

2- KDD (Knowledge Discovery in Databases) Process کے ہر step کی وضاحت کریں۔

3- OLAP اور OLTP کے درمیان فرق بیان کریں۔

4- ایک Classifier کی performance کو evaluate کرنے کے طریقے کیا ہیں؟

5-

a- دیے گئے item set i = {A, B, C, D} کے لیے lattice diagram بنائیں۔

b- دیے گئے support کی بنیاد پر association rules کے لیے confidence calculate کریں۔

6- کسی بھی 3 discretization techniques کی وضاحت کریں۔

7- درج ذیل پر نوٹ لکھیں:

i- Histogram

ii- Quantile Plot

iii- Scatter Plot

8- درج ذیل vectors کے بارے میں اپنی رائے دیں کہ وہ کتنے similar ہیں:

X = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)

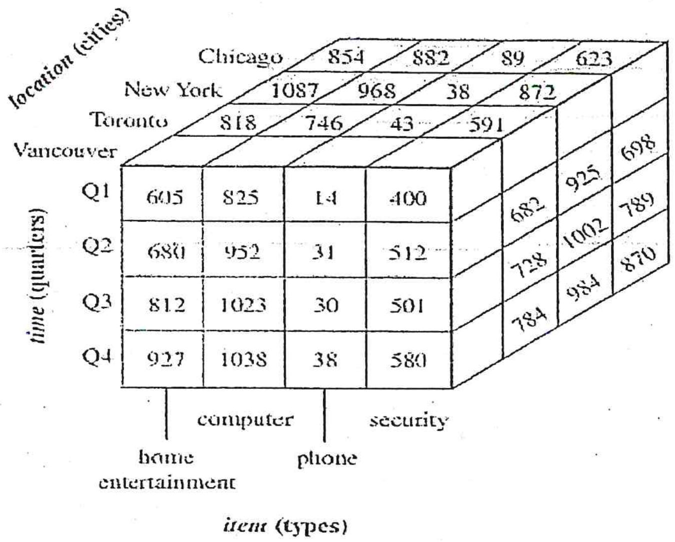
Y = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)

9- Single Linkage Clustering کے concept کی وضاحت کریں۔

حصہ سوم

10- DPP (Data Preprocessing Process) کے مراحل (steps) کو واضح طور پر لکھیں۔

11۔ دیے گئے multidimensional data پر تمام OLAP operations کا اطلاق کریں۔



12۔ دیے گئے dataset کے لیے ایک Decision Tree اخذ (derive) کریں۔

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

13۔ دیے گئے transactional data کے لیے frequent item set معلوم کریں۔

Tid	Items
1	Bread ,Milk
2	Bread,diaper,beer,eggs
3	Milk ,diaper,beer,coke
4	Bread ,milk,diaper,beer
5	Bread, milk,diaper,coke

14۔ دیے گئے data objects کو 3 clusters میں تقسیم کرنے کے لیے K-means Algorithm لکھیں اور اس کا اطلاق کریں۔ ابتدائی طور پر points کو clusters میں اس طرح assign کیا گیا ہے:

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

Object	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