

INFORMATION TECHNOLOGY & DEPARTMENT OF COMPUTER SCIENCE

Programme : Bachelor of Technology (B.Tech)

VI Semester Examination, July 2023

Paper : BTCS611PET : Data Mining and Data Warehousing

Max Marks 70

Time 3 Hours

ہدایات: یہ پرچہ سوالات تین حصوں پر مشتمل ہے:۔ ہر جواب کے لیے لفظوں کی تعداد اشارہ ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

- حصہ اول میں 10 معروضی سوالات ہیں۔ ہر سوال کا جواب لازمی ہے۔ ایک جواب پر 1 نمبر مختص ہے۔ (10x1=10 marks)
- حصہ دوم میں 08 سوالات ہیں۔ ان میں سے طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً (200) لفظوں میں مطلوب ہے۔ ہر جواب پر 6 نمبرات مختص ہیں۔ (5x6=30 marks)
- حصہ سوم میں 5 سوالات ہیں۔ ان میں سے طالب علم کو کوئی تین سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً (500) لفظوں پر مشتمل ہو۔ ہر جواب پر 10 نمبرات مختص ہیں۔ (3x10=30 marks)

حصہ اول

سوال نمبر 1:

- Data Mining کی وضاحت کریں۔
- IQR کی وضاحت کریں۔
- negatively skewed data کا خاکہ بنائیے۔
- binary attributes کے لئے contingency table لکھیں۔
- مندرجہ ذیل attribute type کا نام دیں:
 - Time-to-complete- a task
 - Occupation
 - Eye-color
 - Gender
- Jacquard coefficient کیا ہے۔
- Data Matrix اور Dissimilarity Matrix کے درمیان فرق کریں۔
- Outliers سے noise data کیسے مختلف ہیں۔
- Classifier کی وضاحت کریں۔
- Association Analysis کے Applications کے کچھ کی فہرست بنائیں۔

Section-B

2. صاف ستھرا خاکہ کی مدد سے KDD کے steps کی وضاحت کریں۔
3. i. Data Warehouse کیا ہے۔
ii. OLTP کی وضاحت کریں
4. i. Support اور Confidence کی وضاحت کریں۔
ii. Draw the lattice دیے گئے item set={A,B,C,D} کے لیے
5. Classifier Accuracy کو کیسے evaluate کریں گے۔
6. i. k-means Algorithm لکھیں۔
ii. Data Mining کی دوسری Methodologies کون کون سی ہیں۔
7. Box Plot for the given data and visualize an outlier by using Five number Summary. 7
کے لیے خاکہ بنا دے
18,34,76,29,15,41,46,25,54,38,70,32,43,22
8. Observe the given term frequency vector اور معلوم کریں کہ Y اور X کتنی similar ہیں؛
Give your observation in terms of similarity.
 $X = (1, 3, 4, 0, 0, 4, 2, 8, 7)$
 $Y = (0, 5, 0, 0, 5, 14, 6, 0, 0)$
9. i. فرض کریں۔ maximum اور minimum
values for the attribute income are \$12,000 and \$98,000, respectively, map income to the range [0.0, 1.0]. A value of \$73,600 for income is transformed to which value by using normalization technique.
ii. بحث کریں۔ The various factors affecting the quality of Data.

Section-C

10. Data cleaning اور Data Transformation میں Data pre processing کے مختلف تکنیکوں کی وضاحت کریں۔
11. i. 3-teir architecture of a Data warehouse کا خاکہ بنا کر explain کریں
ii. Data warehouse کے major Features کیا ہیں۔

12. مندرجہ ذیل Observe کریں

transactional database D and find the frequent item set by using Apriori algorithm with Apriori principle

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

13. تعارف لکھتے ہوئے بیان کریں کی-

Applications of Classification Analysis and how to evaluate a classification model is good or not

14. K-means algorithm کو دیے گئے 3 clusters پر Apply کریں ساتھ ہی

Initial points are assigned to clusters as $C1=\{A1,A2,A3\}$ $C2=\{A4,A5,A6\}$ $C3=\{A7,A8\}$

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