

Maulana Azad National Urdu University

M.C.A : IV Semester Examination, June 2021

Paper : MMCA402PCT : Data Warehouse and Data Mining

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 Warehouse ایک Database سے کیسے مختلف ہے؟ اور یہ دونوں کیسے ایک دوسرے کیسے ملتے جلتے (Similar) ہیں؟
2. دیے گئے Data Mining Functionalities کی وضاحت کیجیے۔
- Correlation Analysis ' Characterization ' Association ' Discrimination ' Classification
Outlier Analysis اور Regression, Clustering
3. Support Vector Machines کیا ہوتے ہیں؟ یہ Data کو Classify کرنے میں کس طرح مددگار ہوتے ہیں؟
4. Prediction اور Classification میں ہونے والے مشکلات (Issues) بیان کیجیے۔
5. Data Preprocessing کسے کہتے ہیں؟ Data Preprocessing کی ضروریات (Needs) کے بارے میں سمجھائیے۔
6. Clustering کسے کہتے ہیں؟ Cluster Analysis کی ضروریات بیان کیجیے۔
7. Overfitting کسے کہتے ہیں؟ اس کو کیسے Avoid کیا جاسکتا ہے؟
8. Clustering میں Partitioning Algorithm کے بارے میں سمجھائیے۔
9. Spatial Data Mining اور Multimedia Data Mining کے بارے میں لکھیے۔
10. Association Rule Mining کسے کہتے ہیں؟ یہ Data کو Classify کرنے میں کیسے مدد کرتا ہے؟

حصہ دوم

11. Apriori Algorithm کے بارے میں آپ کیا جانتے ہیں؟ Support Cont 2 کو لیتے ہوئے Apriori Algorithm استعمال کرتے ہوئے دیے گئے Data کو Classify کیجیے۔

T ₁ D	List of Item_1Ds
T ₁	Milk, Butter, Buttermilk
T ₂	Butter, Curd
T ₃	Butter, Bread
T ₄	Milk Butter, Curd
T ₅	Milk, Bread
T ₆	Butter, Bread
T ₇	Milk, Bread
T ₈	Milk, Butter Bread, Butter Milk
T ₉	Milk, Butter, Bread

12. Entropy اور Informatin Gain کسے کہتے ہیں؟ دیے گئے Table میں Data کو استعمال کرتے ہوئے ایک Decision Tree

تیار کیجیے۔

Instances	A2	A2	A3	Class
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Hot	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

13. Naire Bayesian Classifier کے بارے میں آپ کیا جانتے ہیں؟ {M,H} Spieces Classes کے ہر

Conditional Probabilities کو دیے گئے Data کو استعمال کرتے ہوئے Attribute {color, legs, height, smelley}

معلوم کیجیے۔ اسی Probabilities کو استعمال کرتے ہوئے (color = green, legs = 2, height = tall & smelley = no)

اس کی Probability Values معلوم کیجیے۔

No.	Color	Legs	Height	Smelley	Spieces
1	White	3	Sort	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

.14 K-Medoids Clustering Algorithm کے بارے میں سمجھائیے۔ K-Medoids Algorithm کو استعمال کرتے ہوئے دیے

گئے Datapoints کو 2 Clusters میں Classify کیجیے۔

X	Y
7	6
2	6
3	8
18	5
7	4
4	7
6	2
7	3
6	4
3	4

.15 Web Mining کسے کہتے ہیں؟ مختلف Web Mining کے بارے میں سمجھائیے۔ Web Mining کے Application بیان کیجیے۔

☆☆☆