

Maulana Azad National Urdu University
B.Tech VI Semester Examination - May - 2018
Paper - BTCS604PCT : Data Warehousing & 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 اشارہ ہے۔ تمام حصوں سے سوالوں کا جواب دینا لازمی ہے۔

1. حصہ اول میں 10 لازمی سوالات ہیں جو کہ معروضی سوالات/خالی جگہ پر 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، اور اس میں طالب علم کو کوئی پانچ سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً دو سو (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 میں Noise کس طرح داخل ہوتی ہے؟ دو وجہ بتاؤ۔
- (ii) Dimension Table کیا ہے؟
- (iii) Prediction اور Classification میں فرق بتائیے۔
- (iv) DMQL کا Syntax لکھیے۔
- (v) کسی Attribute کی Entropy کو Calculate کرنے کا Formula لکھیے۔
- (vi) کسی Classifier Model کی Accuracy کس طرح حاصل کی جاتی ہے؟
- (vii) Intra Cluster اور Inter Cluster کے درمیان فرق بتائیے۔
- (viii) Association Rule Mining میں Support کو Define کیجیے۔
- (ix) کوئی 3 Data Mining Tools لکھیے۔
- (x) Pre-Pruning اور Post-Pruning کے Methods کے درمیان فرق بتائیے۔

حصہ دوم

- (2) Data Smoothing کے لیے Binning Method کا استعمال کیا جاتا ہے۔ بتائیے کہ کیا یہ Discretization Method کے لیے کس طرح استعمال کیے جاسکتے ہیں؟
- (3) مندرجہ ذیل کی تعریف کیجیے۔
- (4) OLAP اور OLTP کے درمیان فرق واضح کرو۔ OLAP کے Typical Operations کے بارے میں آپ کیا جانتے ہیں؟ واضح کرو
- (5) تفصیل کے ساتھ سمجھائیے کہ Association Rule Mining یا تو Supervised or Unsupervised Type ہے
- (6) ایک صاف Sketch-diagram کی مدد سے کسی Data Warehouse کے Architecture کو واضح کیجیے۔
- (7) کسی Credit Card Company کے لیے کسی Fraud کو Identify کرنے کے لیے آپ کیا قدم اٹھائیں گے؟
- (8) Classification اور Clustering کے درمیان فرق کو واضح کرو اور کسی Application کا ایک Informal Example دیجیے جو ہر ایک Technique سے وابستہ ہو۔
- (9) Knowledge Discovery in Databases (KDD) میں Data Pre-Processing کا کیا مقصد ہے؟ ہر ایک Component کو ایک مثال کے ساتھ واضح کیجیے۔

حصہ سوم

- (10) Data Mining میں (Market-Basket Analysis) کا کیا مقصد ہے؟ Minimum Support = 55% اور Minimum Confidence 60% کے ساتھ درج ذیل Dataset سے Frequent Item Set حاصل کرنے کے لیے Apriori Algorithm لگائیے۔

Transaction ID	Item Description
T1	Milk, Bread, Butter, Chocolet, Peanut
T2	Bread, Checolet, Peanut, Paneer
T3	Milk, Butter, Peanut, Curd, Eggs, Ice-cream
T4	Bread, Butter, Peanut, Paneer
T5	Butter, Chocolet, Curd, Eggs, Ice-cream
T6	Bread, Butter, Curd, Paneer
T7	Milk, Bread, Butter, Curd, Eggs, Ice-cream
T8	Curd, Eggs, Ice-cream, Butter, Chocolet, Milk
T9	Curd, Eggs, Ice-cream, Butter, Chocolet, Milk,
T10	Curd, Eggs, Ice-cream, Chocolet, Milk

(11) STAR Schema کے Design کے Data Ware House کو Construct کرو۔ ایک مثال کے ساتھ

Snow Flake Schema Data Warehouse Design کے فوائد کو بیان کیجیے۔

(12) درج ذیل Attributes کو Classify کرو جیسے Discrete, Binary یا Continuous۔ اور ان کو Qualitative (Nominal or Ordinal) یا Quantitative (Interval or Ratio) کی طرح بھی Classify کرو۔ کچھ Cases میں ایک سے زیادہ

Interpretation ہو سکتے ہیں۔ اس لیے وجہ (Reason) کو واضح کیجیے۔

Time in terms of AM or PM (a)

Height above sea level (b)

Bronze, Silver and Gold Medals as awarded at the olympics (c)

Ability to pass light in terms of the following values: Opaque translucent, transparent. (d)

Density of a substance in grams per cubic centimeter (e)

Density of a substance in grams per cubic centimeter (e)

(13) درج ذیل Data Set (for playing tennis) کا استعمال کرتے ہوئے Bayesian Classification کے Method کو تفصیل کے

ساتھ بیان کیجیے۔ (Game Playing Weather Problem)

کسی unknown / unseen sample X = <ran, hot, high, false> کو Classify کرو۔

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	p
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

(14) درج ذیل Data Set کا استعمال کرتے ہوئے Decision Tree Induction Method کی Classification Technique کو تفصیل کے ساتھ بیان کرو۔ Tree کو Construct کرنے کے لیے Information Gain کا استعمال کرتے ہوئے کسی Attribute کو Select کرنے کے Process کو دکھاؤ۔ سبھی Computations کو تفصیل سے بیان کیجیے۔

Age_group	Income_Group	Student_Category	Credit_Rating	Buys_Computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

☆☆☆