Maulana Azad National Urdu University

B.Tech VI Semester Examination - May - 2019

Paper - BTCS604PCT : Data Warehousing & Data Mining

Time : 3 hrs **Marks : 70**

ہدایات:

یہ پرچہ سُوالات 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، اور 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 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) اب کس طرح Mining کے لیے Data کی Preprocessing کے دوران Missing/NULL Values کو Treat کروگے؟ کوئی دو Methods کا ذکر کیجیے۔

(ii) Categorical اور Numerical Data کے درمیان فرق کو ایک مثال کی مدد سے بتائے۔

(iii) کوئی 4 Challenges جو کہ Data Mining میں آتے ہیں بتائے۔

(iv) کوئی 3 Typical OLAP Operations لکھیے۔

(v) Noise Data کو کیسے Handle کرتے ہیں؟

(vi) کسی Dimension Table سے آپ کیا سمجھتے ہیں؟

(vii) آپ Classification اور Prediction میں کیسے فرق کریں گے؟

(viii) DMQL کا Syntax لکھیے۔

(ix) کسی Attribute کی Entropy کو Calculate کرنے کے لیے Formula بتایئے۔

(x) ایک Classifier کی Accuracy کو کیسے Measure کیا جاتا ہے؟

حصہ دوم

(2) Database اور Data Warehouse کے درمیان فرق Online Transaction Processing اور Online Analytical Processing کے Context میں کریں۔

(3) ایک مناسب خاکے کی مدد سے Database میں Knowledge Discovery کے دوران Data Mining میں آنے والے Phases کی تعریف کیجیے۔

(4) درج ذیل کی تعریف بیان کیجیے۔

(a) Frequent Sets (b) Confidence (c) Support (d) Association Rule

(5) ایک صاف خاکے کی مدد سے Data Warehouse کے Architecture کی تفصیلی وضاحت کیجیے۔

(6) Data Mining کے مقصد کے لیے Data کے Collection اور Availability کے Major Source کی تفصیلی وضاحت کیجیے۔

(7) ایک مناسب مثال کی مدد سے Partitional Clustering کی تفصیلی وضاحت کیجیے۔

(8) ایک مناسب مثال کی مدد سے Star Schema اور Snowflake Schema کے درمیان فرق کو واضح کریں۔

(9) Discretization اور Concept Hierarchy کا کیا مقصد ہے؟

حصہ سوم

(10) Data Mining میں Market-Basket Analyses کا کیا مقصد ہے؟ درجہ ذیل Dataset جس کے لیے Minimum Support 55% اور Minimum Confidence = 60% ہے سبھی Steps ظاہر کرتے ہوئے Apiori Algorithm کی مدد سے Frequent Item-Set حاصل کریں۔

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) Data Warehouse Design کے لیے STAR Schema کو بنائے اور Data Warehouse Design میں Snowflake Schema کے فوائد کو مناسب مثال کی مدد سے واضح طور پر سمجھائیے۔

(12) درجہ ذیل Attributes کو Binary, Discrete or Continuous میں تقسیم کیجیے۔ ان کو Qualitative یا Quantitative میں بھی تقسیم کرئے۔ کچھ Cases میں سے ایک سے زیادہ Interpretation ہوسکتا ہے، اس کو صحیح سے واضح کیجیے۔

(a) Time in Terms of AM or PM (b) Height above sea Level

(c) Olympics کے دوران حاصل کیے گئے Bronze, Silver اور Gold Medals

(d) Opaque, Translucent اور Transparent کی Terms میں Light کو Pass کرنے کی Ability

(13) Teniss کے لیے درجہ ذیل Dataset کا استعمال کرتے ہوئے Bayesian Classification Method کو واضح کرئے۔ کسی *Unknown / Unseen sample X = < ran, hot, high, false >* کو Classify کرے۔

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) Decision Tree Induction Method کا استعمال کرتے ہوئے درجہ ذیل Dataset کے لیے Classification Technique کو واضح کرئے۔ Tree کو بنانے کے لیے Information Gain کا استعمال کرتے ہوئے کسی Attribute کے Selection کے Process کو دکھایئے۔ Computations کو صحیح سے ظاہر کریں۔

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

☆☆☆