

Maulana Azad National Urdu University

B.Tech VI Semester Examination - May - 2017

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) KDD (Process) عمل میں ہونے والے مختلف (Steps) مراحل کیا ہیں۔
- (ii) کوئی بھی چند Data Mining Techniques (Some) بیان کیجئے۔
- (iii) Predictive Model ماڈل کی وضاحت کیجئے۔
- (iv) Outlier Mining کسے کہتے ہیں؟
- (v) Data Warehouse کی وضاحت کیجئے۔
- (vi) Operational Databases کیا ہیں۔
- (vii) OLAP System میں Data Base Design کو کیسے Represent کیا جاتا ہے۔
- (viii) Dimension Table کی وضاحت کیجئے۔
- (ix) Star Schema اور Snowflake Schema ماڈل کے بیچ اہم فرق واضح کیجئے (Major Difference)۔
- (x) Multidimensional Data Model میں OLAP Operations بیان کیجئے۔

حصہ دوم

- (2) Knowledge Discovery اور Data Mining کے بیچ فرق (Differences) سمجھائیے۔
- (3) ایک Bank Manager کے لحاظ سے آپ یہ کیسے طے کریں گے کہ آپ کو ایک Applicant کو Loan دینا چاہیے یا نہیں۔ Data Mining Techniques کے ذریعہ اس کا خلاصہ (Elaborate) کیجیے۔
- (4) ایک Credit Card Company کے (Frauds) (دھوکے بازی) کے بارے میں جاننے کے لیے Data Mining Functionalities کی مدد سے آپ کیا (Steps) اقدامات لیگیں۔
- (5) کسی بھی تین اہم Applications (Potential) سمجھائیے جس کے لیے Data Mining Techniques مناسب (Appropriate) ہو۔
- (6) Classification اور Clustering کے بیچ فرق (Difference) واضح کیجیے اور ایک Application کی مثال (Informal Example) دیکھیے جو ان Techniques سے فائدہ مند ہو۔
- (7) Snowflake Schema اور Star Schema سے کس طرح مختلف ہے Snowflake Schema اور Star Schema کے دونوں (Advantages) اور (Dis-advantages) نقصان بیان کیجیے۔
- (8) Clustering کے دو طریقوں کو سمجھائیے۔
- (a) Grid Based Clustering
- (b) Model Based Clustering
- (c) Density Based Clustering
- (9) Data Warehouse ایک Data Base سے کس طرح مختلف ہے (Different) ہے اور وہ کیسے ملتے جلتے (Similar) ہیں۔

حصہ سوم

- (10) Data Mining میں Market - Basket Analysis کا کیا (Purpose) مقصد ہے۔ دیے گئے Data Set کو دیکھیے جس میں Min Support - 55% ہے اور Minimum Confidence - 60% ہے۔ Apriori Algorithm کو استعمال کرتے ہوئے Frequent Item - Set معلوم کیجیے۔ ہر مراحل (Steps) کو اچھی طرح سے سمجھائیے۔

Transaction Id	Item Description
T1	Milk, Bread, Butter, Choclet, Peanut
T2	Bread, Choclet, Peanut, Paneer
T3	Milk, Butter, Peanut, Curd, Eggs, Ice Cream
T4	Bread, Butter, Peanut, Paneer
T5	Butter, Choclet, Curd, Eggs, Ice Cream
T6	Bread, Butter, Curd, Paneer
T7	Milk, Bread, Butter, Curd, Eggs, Ice Cream
T8	Curd, Eggs, Ice Cream, Butter, Choclet, Milk
T9	Curd, Eggs, Ice Cream, Butter, Choclet, Mild
T10	Cure, Eggs, Ice Cream, Choclet, Mild

(11) مندرجہ ذیل کے بارے میں مختصر نوٹ لکھیے۔

Non - Linear Regression (a)

Spatial Data Mining (b)

Multimedia Data Mining (c)

Generalization (d)

(12) Over Fitting کی Working Principle کارکردگی کے اصول سمجھائیے۔ (SVM) Support Vector Machine

SVM سے کیسے (Avoids) بچتا ہے۔

(13) Databases میں Knowledge Discovery (KDD) میں Data - Processing کا کیا مقصد ہے۔ ہر

Component کو اپنی پسند کی مثال کے ذریعہ سمجھائیے۔

(14) Bayesian Classification کے لئے کھیلنے کے Data - Set کے ذریعہ (Tennis) Playing کو سمجھائیے۔

Method کو سمجھائیے۔

Sample X = <rain, hot, high, false> unknown / an unseen کا خلاصہ کیجیے۔

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

☆☆☆