

مولانا آزاد نیشنل اردو یونیورسٹی

Diploma in Engineering

VI Semester Exams: AICTE - April 2024

Code: DPCS616PET Subject: Data Science: DW and DM

Total Time: 3Hr

Total Marks: 60

ہدایات:

یہ پرچہ سوالات پر مشتمل ہے: حصہ اول، حصہ دوم اور 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 / حتمی جگہ پر کرنا / مختصر جواب والے سوالات ہیں۔ ہر سوال کا جواب لازمی ہے۔ ہر سوال کے لئے 1 نمبر مختص ہے۔

2. (10x1=10) حصہ دوم میں 7 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 4 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً دو سو (200) لفظوں پر مشتمل ہو۔ ہر سوال کے لئے 5 نمبرات مختص ہیں۔ (4x5=20)

3. حصہ سوم میں 5 سوالات ہیں۔ اس میں سے طالب علم کو کوئی 3 سوالوں کے جواب دینے ہیں۔ ہر سوال کا جواب تقریباً پانچ سو (500) لفظوں پر مشتمل ہو۔ ہر سوال کے لئے 10 نمبرات مختص ہیں۔ (3x10=30)

حصہ اول

1. i- Data _____ Enterprise کو میموری فراہم کرتا ہے۔

ii- DMQL کا مخفف کیا ہے؟

iii- ڈیٹا کے معیار کا کنٹریبوشن جہتی پیمانہ (multi-dimensional measure of data quality) لکھیں۔

iv- (Data Mining) ڈیٹا مائننگ انٹرپرائز Enterprise کو _____ فراہم کرتی ہے

v- فلیٹ فائل (flat file) میں کس قسم کا ڈیٹا پایا جاتا ہے؟

vi- Centralized data warehouse بنائیں۔

vii- Query-Driven Approach ڈرا کریں۔

viii- metadata میں fields کی کتنی اقسام پائی جاتی ہیں؟ لکھیں۔

ix-missing data ہونے کی دو وجوہات لکھیں۔

x-Federated data warehouse ڈرا کریں۔

حصہ دوم

2- data pre-processing میں اہم کام (Major task in data preprocessing) کی وضاحت کریں۔

3- Classification کیا ہے؟ training اور testing کے ڈیٹا کے ساتھ Classification کا مظاہرہ کریں۔

4- درج ذیل پیرامیٹرز پر غور کر کے OLTP اور OLAP کے درمیان فرق لکھیں۔

Users, functions, database design, data, usage, access, #records access, #users

5- Hierarchical اور density clustering کے طریقوں کی وضاحت کریں؟

6- Prediction سے آپ کا کیا مطلب ہے؟ Predictive modeling in databases کی وضاحت کریں۔

7- data warehouse سے آپ کا کیا مطلب ہے؟ heterogeneous database اور data warehouse کے درمیان

فرق لکھیں۔

8- Various types of association rule کی وضاحت کریں؟

حصہ سوم

9- ایک مناسب خاکہ کی مدد سے، knowledge discovery in databases (KDD) میں ڈیٹا

مانٹنگ (data mining) کے process کی وضاحت کریں۔

10- سپورٹ ویکٹر مشین (Support Vector Machine) کیا ہے؟ مثال کے ساتھ مظاہرہ کریں، Linear

اور SVM non-linear کی تعریف کریں۔

11- درج ذیل پر روشنی ڈالیں:

(a)- ڈیٹا سائنس (Data Science)۔

(b)- مشین لرننگ (Machine Learning)۔

(c)- بڑا ڈیٹا (Big Data)۔

12- decision tree کی اس کے phases کے ساتھ تعریف کریں۔ مندرجہ ذیل ٹریننگ سیٹ کا decision

tree بنائیں۔

Training Dataset

age	income	student	credit_rating
<=30	high	no	fair
<=30	high	no	excellent
31...40	high	no	fair
>40	medium	no	fair
>40	low	yes	fair
>40	low	yes	excellent
31...40	low	yes	excellent
<=30	medium	no	fair
<=30	low	yes	fair
>40	medium	yes	fair
<=30	medium	yes	excellent
31...40	medium	no	excellent
31...40	high	yes	fair
>40	medium	no	excellent

Two popularly used approaches handle noisy data کے دوران data pre-processing -13

استعمال کیے جاتے ہیں، binning method اور clustering method۔ کسی مناسب مثال کے ساتھ دونوں techniques کی وضاحت کریں۔
