

**GUJARAT TECHNOLOGICAL UNIVERSITY****DIPLOMA ENGINEERING – SEMESTER –6 (NEW)- EXAMINATION –WINTER-2022****Subject Code: 3361604****Date: 15-12-2022****Subject Name: Data Mining And Warehousing****Time: 02:30 PM TO 05:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Use of simple calculators and non-programmable scientific calculators are permitted.
1. English version is authentic.

- Q.1** Answer any seven out of ten. દશમાંથી કોઈપણ સાતના જવાબ આપો. **14**
1. What is a data mining?  
૧. ડેટા માઈનિંગ શું છે?
  2. List any four applications of data mining.  
૨. ડેટા માઈનિંગની કોઈપણ ચાર એપ્લિકેશનની યાદી બનાવો.
  3. Define continuous attributes in data mining.  
૩. ડેટા માઈનિંગમાં continuous attributes ને વ્યાખ્યાયિત કરો.
  4. List major tasks involved in data preprocessing.  
૪. ડેટા પ્રીપ્રોસેસિંગમાં સામેલ મુખ્ય કાર્યોની યાદી બનાવો.
  5. Write full form of ARFF and WEKA.  
૫. ARFF અને WEKA નું પૂર્ણ સ્વરૂપ લખો.
  6. What is clustering in data mining?  
૬. ડેટા માઈનિંગમાં ક્લસ્ટરિંગ શું છે?
  7. What is redundancy in data integration?  
૭. ડેટા એકીકરણમાં રીડન્ડન્સી શું છે?
  8. What is classification in data mining?  
૮. ડેટા માઈનિંગમાં વર્ગીકરણ શું છે?
  9. What is data mart?  
૯. ડેટા માર્ટ શું છે?
  10. What is WEKA?  
૧૦. WEKA શું છે?
- Q.2** (a) What are the different techniques of data mining? Explain in brief. **03**  
પ્રશ્ન. ૨ (અ) ડેટા માઈનિંગની વિવિધ તકનીકો શું છે? ટૂંકમાં સમજાવો. **૦૩**
- OR
- (a) Briefly explain types of data in data mining. **03**  
(અ) ડેટા માઈનિંગમાં ડેટાના પ્રકારોને સંક્ષિપ્તમાં સમજાવો. **૦૩**
- (b) What are the major applications of data mining? Explain in brief. **03**  
(બ) ડેટા માઈનિંગના મુખ્ય એપ્લિકેશનો શું છે? ટૂંકમાં સમજાવો. **૦૩**
- OR
- (b) Briefly explain challenges of data mining. **03**  
(બ) ડેટા માઈનિંગના પડકારોને સંક્ષિપ્તમાં સમજાવો. **૦૩**

- (c) What is standard deviation and variance in statistical calculations? Explain with example. **04**
- (ક) આંકડાકીય ગણતરીઓમાં standard deviation અને variance શું છે? ઉદાહરણ સાથે સમજાવો. **૦૪**

OR

- (c) What is range and quartile in statistical calculations? Explain with example. **04**
- (ક) આંકડાકીય ગણતરીમાં range અને quartile શું છે? ઉદાહરણ સાથે સમજાવો. **૦૪**
- (d) Explain nominal and binary attributes with example. **04**
- (ડ) nominal અને binary એટ્રિબ્યુટ્સ ઉદાહરણ સાથે સમજાવો. **૦૪**

OR

- (d) Explain ordinal and numeric attributes with example. **04**
- (ડ) ordinal અને numeric એટ્રિબ્યુટ્સ ઉદાહરણ સાથે સમજાવો. **૦૪**

**Q.3**  
**પ્રશ્ન. ૩**

- (a) What is data cleaning and why is it important? **03**
- (અ) ડેટા ક્લિનિંગ શું છે અને તે શા માટે મહત્વપૂર્ણ છે? **૦૩**

OR

- (a) What is correlation analysis? How is correlation analysis used to detect redundancies? **03**
- (અ) correlation analysis શું છે? રિડન્ડન્સી શોધવા માટે correlation analysis નો ઉપયોગ કેવી રીતે થાય છે? **૦૩**
- (b) How do we perform data visualization in WEKA? Explain in brief. **03**
- (બ) આપણે WEKA માં ડેટા વિઝ્યુલાઇઝેશન કેવી રીતે કરી શકીએ? ટૂંકમાં સમજાવો. **૦૩**

OR

- (b) How do we perform data filtering in WEKA? Explain in brief. **03**
- (બ) આપણે WEKA માં ડેટા ફિલ્ટરિંગ કેવી રીતે કરી શકીએ? ટૂંકમાં સમજાવો. **૦૩**
- (c) Explain missing values and noisy data in data cleaning. **04**
- (ક) ડેટા ક્લિનિંગમાં missing values અને noisy data સમજાવો. **૦૪**

OR

- (c) Explain tuple duplication and identification problem in data integration. **04**
- (ક) ડેટા એકીકરણમાં ટપલ ડુપ્લિકેશન અને ઓળખની સમસ્યા સમજાવો. **૦૪**
- (d) Explain the concept of probability-based solving. **04**
- (ડ) probability-based solving નો concept સમજાવો. **૦૪**

OR

- (d) What is clustering? Explain its types. **04**
- (ડ) ક્લસ્ટરિંગ શું છે? તેના પ્રકારો સમજાવો. **૦૪**

**Q.4**  
**પ્રશ્ન. ૪**

- (a) What is a data warehouse? What are the benefits of a data warehouse? **03**
- (અ) ડેટા વેરહાઉસ શું છે? ડેટા વેરહાઉસના ફાયદા શું છે? **૦૩**

OR

- (a) Describe any three differences between operational database systems and data warehouse. **03**
- (અ) ઓપરેશનલ ડેટાબેઝ સિસ્ટમ્સ અને ડેટા વેરહાઉસ વચ્ચેના કોઈપણ ત્રણ તફાવતોનું વર્ણન કરો. **૦૩**
- (b) Explain enterprise data warehouse model. **04**
- (બ) એન્ટરપ્રાઇઝ ડેટા વેરહાઉસ મોડલ સમજાવો. **૦૪**

OR

- (b) Explain virtual data warehouse model. **04**

	(બ) વર્ચ્યુઅલ ડેટા વેરહાઉસ મોડલ સમજાવો.	૦૪
	(c) What is decision tree in data mining? Explain ID3 decision tree algorithm with example.	07
	(ક) ડેટા માઈનિંગમાં decision tree શું છે? ઉદાહરણ સાથે ID3 decision tree અલ્ગોરિધમ સમજાવો.	૦૭
<b>Q.5</b>	(a) How does the data warehousing and data mining work together?	<b>04</b>
<b>પ્રશ્ન. ૫</b>	(અ) ડેટા વેરહાઉસિંગ અને ડેટા માઈનિંગ એકસાથે કેવી રીતે કાર્ય કરે છે?	૦૪
	(b) What is the difference between OLAP and OLTP?	<b>04</b>
	(બ) OLAP અને OLTP વચ્ચે શું તફાવત છે?	૦૪
	(c) Define following terms. 1. Mean 2. Median 3. Mode	<b>03</b>
	(ક) નીચેના શબ્દો વ્યાખ્યાયિત કરો. 1. Mean 2. Median 3. Mode	૦૩
	(d) What is ARFF file in WEKA? Can we use csv file in WEKA?	<b>03</b>
	(ડ) WEKA માં ARFF ફાઈલ શું છે? શું આપણે WEKA માં csv ફાઈલનો ઉપયોગ કરી શકીએ?	૦૩

\*\*\*\*\*