

**B.C.S.E. FINAL EXAMINATION, 2018**  
**(1<sup>st</sup> SEMESTER)**  
**DATA MINING**

Time : Three Hours

Full Marks 100

Answer Question 1 and any 4 questions from the rest

All questions carry equal marks

Any subpart of a question carry equal marks unless otherwise specified

Q1. Answer briefly with justification or examples as and when necessary :-

- (i) Mention how the following are similar or dissimilar
- (a) PCA
  - (b) DWT
  - (c) histograms
  - (d) data cube aggregation
- (ii) How does the Jaccard coefficient measure similarity ?
- (iii) Which one is a better correlation measure and why ?
- (a) lift
  - (b) cosine
- (iv) Which one gives the best central tendency for skewed data and why ?
- (a) Mean
  - (b) Trimmed mean
  - (c) Median
  - (d) Mode
- (v) What is the difference between noise and outlier?

Q2. Measure correlation of Students taking up Machine Learning and Cryptography courses, as per the Contingency Table given below, using both the datasets and techniques mentioned in the following Data-cum-Correlation Table, thus populating the tables properly :-

Contingency Table

	Machine Learning	No Machine Learning	Erow
Cryptography	MC	M'C	C
No Cryptography	MC'	M'C'	C'
Ecol	M	M'	E

Data-cum-Correlation Table

Data Set	MC	M'C	MC'	M'C'	Chi-Square	All-Conf
D1	10,000	1000	1000	10,000		
D2	10,000	1000	1000	100		

Draw your conclusions regarding the efficacy of both measures.

- Q3. (a) Distinguish between the terms OLTP and OLAP on the basis of some general features. 7
- (b) Explain the multi-dimensional Data Model in detail. 7
- (c) What are the different Warehouse Models ? 6

[contd..]

- Q4. (a) Explain the terms :- Association Rules, Support, Confidence 6  
(b) What is the Apriori Method ? 7  
(c) What are its disadvantages ? 3  
(d) Mention a technique to overcome the same. 4
- Q5. (a) What are supervised, semi-supervised and unsupervised modes of learning?  
(b) Explain how a Multilayer Feed-Forward Neural Network learns with the help of the Backpropagation Algorithm. 6 + 14
- Q6. (a) Discuss the role of clustering in data mining. 6  
(b) Elaborate on a basic partitioning method. 14
- Q7. Describe any two of the following in detail :  
(a) Proximity Measures based on different types of attribute  
(b) Fuzzy Clustering using EM algorithm  
(c) Categorizations of Outlier Detection methods  
(d) Data Mining techniques on Sequence Data