B.Tech. Degree VI Semester Special Supplementary Examination July 2019

CS 15-1604 DATA MINING
(2015 Scheme)

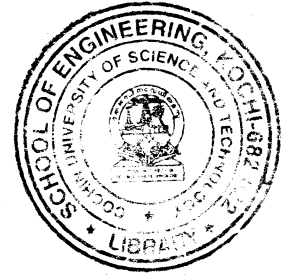Time: 3 Hours   Maximum Marks: 60

PART A
(Answer *ALL* questions)

(10 × 2 = 20)

I. (a) Distinguish Data integration and normalization?

(b) Distinguish supervised and unsupervised learning with examples?

(c) Explain Basic OLAP operations on cubes?

(d) Explain SVM's with an example?

(e) Explain support and confidence in association rules mining?

(f) What is density based clustering? explain with example?

(g) Write a note on linear regression?

(h) What are 3V's of big data?

(i) Explain the hadoop distributed architecture?

(j) Write notes on Hive and HBase?

PART B

(4 × 10 = 40)

II. Explain the various applications of data mining, namely time series and   (10)
clustering, with examples for each?

OR

III. Explain the various stages in data mining with diagram and examples?   (10)

IV. Explain any attribute dependency/ feature selection methods such as chi-   (10)
square analysis in data mining with examples?

OR

V. Consider the following dataset   (10)

1	{Apple, bread, cake, dates}
2	{ Apple, bread}
3	{bread, cake, dates}
4	{ banana, cake }
5	{ Apple, banana, dates }
6	{ cake, dates }
7	{ banana, dates }

Assume that the minimum support expected is 3/7 and minimum
confidence expected is 3/4. Compute the association rules following
apriori?

(P.T.O.)

| VI. | (a) | Explain any one Heirarchical based clustering algorithm?Explain both bottom up and top down algorithms? | (5) |
| | (b) | What are different types of clustering? Explain the context of each of them? | (5) |

OR

| VII. | | Distinguish K-Means and KNN with examples? | (10) |

| VIII. | | Explain the ecosystem in hadoop including diagrams and examples? | (10) |

OR

| IX. | | What is the mapreduce framework? How you can sort 10000 student records using total marks given marks in different subjects using the mapreduce concept? | (10) |

***