

**Data Mining & Analytics (PCC-DS-303G)**

**Assignment 2**

1. a) Define support and confidence.  
b) What are the various tasks in frequent pattern mining?  
c) What is the need of Associative Mining?  
d) What are the limitations of Apriori Algorithm?  
e) What are advantages of correlation analysis?
2. a) What is Market Basket Analysis. Explain it with the help of suitable example.  
b) Compare and contrast various pattern evaluation measures.
3. Apply Apriori algorithm to find frequent item sets and also generate association rules for the following data

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

4. How to improve the efficiency of Apriori based mining? Explain any two methods to improve it.
5. Apply Frequent Pattern Growth algorithm to the following data

Transaction ID	Items
T1	{E, K, M, N, O, Y}
T2	{D, E, K, N, O, Y}
T3	{A, E, K, M}
T4	{C, K, M, U, Y}
T5	{C, E, I, K, O, O}

6. Write algorithm for mining frequent item sets using vertical data format.
- 7.

Given the following pairs of values:

Capital Employed (Rs. In Crore)	1	2	3	4	5	7	8	9	11	12
Profit (Rs. In Lakhs)	3	5	4	7	9	8	10	11	12	14

(a) Draw a scatter diagram

(b) Do you think that there is any correlation between profits and capital employed? Is it positive or negative? Is it high or low?

8.

From following information find the correlation coefficient between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method.

Firm	1	2	3	4	5	6	7	8	9	10
Advertisement Exp. (Rs. In Lakhs)	11	13	14	16	16	15	15	14	13	13
Sales Volume (Rs. In Lakhs)	50	50	55	60	65	65	65	60	60	50

9.

Find the correlation coefficient between age and playing habits of the following students using Karl Pearson's coefficient of correlation method.

Age	15	16	17	18	19	20
Number of students	250	200	150	120	100	80
Regular Players	200	150	90	48	30	12

10.

A computer while calculating the correlation coefficient between the variable X and Y obtained the following results:

$$N = 30; \quad \sum X = 120 \quad \sum X^2 = 600 \quad \sum Y = 90 \quad \sum Y^2 = 250 \quad \sum XY = 335$$

It was, however, later discovered at the time of checking that it had copied down two pairs of observations as: (X, Y) : (8, 10) (12, 7)

While the correct values were: (X, Y) : (8, 12) (10, 8)

Obtain the correct value of the correlation coefficient between X and Y.

11.

Find out spearman's coefficient of correlation between the two kinds of assessment of graduate students' performance in a college.

Name of students	A	B	C	D	E	F	G	H	I
Internal Exam	51	68	73	46	50	65	47	38	60
External Exam	49	72	74	44	58	66	50	30	35

12.

From the following data, compute the rank correlation.

X	82	68	75	61	68	73	85	68
Y	81	71	71	68	62	69	80	70