87-mk : 1stHf-11.

Con. 2937-11.        (REVISED COURSE)        **RK-3498**

(3 Hours)        [ Total Marks : 100

N.B. (1) Question No. 1 is **compulsory**.
      (2) Attempt any **four** questions from remaining **six** questions.
      (3) Assume **suitable** data if **required**.

1. (a) Explain BIRCH with example.      5
    (b) Write FP-growth Algorithm.      5
    (c) Define and explain :      5
       (i) Support, (ii) Confidence, (iii) Information Gain, (iv) Entropy,
       (v) Gini Index.
    (d) Explain Web Content Mining and Web Usage Mining.      5

2. (a) Write difference between OLTP and OLAP. Also explain different OLAP operations    10
       and applications.
    (b) Apply the Apriori Algorithm on the following data with Minimum Support = 2.    10

| TID | List of item IDS |
|------|------------------|
| T100 | I1, I2, I4 |
| T200 | I1, I2, I5 |
| T300 | I1, I3, I5 |
| T400 | I2, I4 |
| T500 | I2, I3 |
| T600 | I1, I2, I3, I5 |
| T700 | I1, I3 |
| T800 | I1, I2, I3 |
| T900 | I2, I3 |
| T1000 | I3, I5 |

3. (a) Explain KDD process and architecture of Typical Data Mining System.    10
    (b) Predict a class label of an unknown tuple X = { age = '< = 20', Income = 'Medium',    10
       Student = 'Yes', Credit rating = 'Fair' } using Naive Bayesian classification.

| Age | Income | Student | Credit_rating | Class: buys_Laptop |
|------|--------|---------|---------------|---------------------|
| >30 | Medium | No | excellent | No |
| < = 20 | High | No | Fair | No |
| 21 - 3 | High | Yes | Fair | Yes |
| < = 20 | High | No | excellent | No |
| 21 - 30 | Medium | No | excellent | Yes |
| 21 - 30 | High | No | Fair | yes |
| < = 20 | Medium | Yes | excellent | Yes |
| >30 | Medium | No | Fair | Yes |
| >30 | Medium | Yes | Fair | Yes |
| >30 | Low | Yes | Fair | Yes |
| < = 20 | Low | Yes | Fair | Yes |
| >30 | Low | Yes | excellent | No |
| 21 - 30 | Low | Yes | excellent | Yes |
| < = 30 | Medium | No | Fair | No |

[ TURN OVER

4. (a) Explain dimensionality reduction for text. Also explain different text mining 10
      approaches.
   (b) Use K-means algorithm to create 3 clusters for given set of values ( 2, 3, 6, 8, 5
      9, 12, 15, 18, 22 ) .
   (c) Explain Agglomerative clustering with example. 5

5. (a) Explain Data Cleaning, Data Integration and Transformation in detail. 10
   (b) What is Stream Data ? Explain Hoeffding Tree Algorithm with example. 10

6. (a) Explain Data Mining for Market segmentation and retail industry. 10
   (b) Explain different Data Reduction techniques. 10

7. Write short notes on (any **four**) :— 20
      (a) Market Basket Analysis
      (b) Spatial Data Cube Construction
      (c) Linear Regression
      (d) Constraint-Based Association Mining
      (e) K-Mediods.