(2 ½ Hours) **[Total Marks: 75]**

N.B. 1) All questions are compulsory.
2) Figures to the right indicate marks.
3) Illustrations, in-depth answers and diagrams will be appreciated.
4) Mixing of sub-questions is not allowed.

Q. 1 Attempt ANY FOUR from the following: **(20M)**

(a) Elaborate features of Data Warehouse.
(b) Explain Data cube in brief.
(c) What do you mean by many to many relationship in context with Data Warehouse?
(d) Write a note on Snowflake schema.
(e) Discuss hybrid and virtual types of Data Warehouse.
(f) Define functional dependencies of data.

Q. 2 Attempt ANY FOUR from the following: **(20M)**

(a) Explain process of KDD in detail.
(b) Explain Apriori algorithm with proper example.
(c) Construct FP growth tree for the following transactions. (min support= 3)

TID	Items
T1	m,a,b,s,d,e,f
T2	a, b, s
T3	s, b,m
T4	m,d,e,a
T5	d,e,f,s
T6	b,s,a,d,f

(d) Discuss various data mining techniques.
(e) Write a note on process of data cleaning with proper example.
(f) What are social implication of data mining?

Q. 3 Attempt ANY FOUR from the following: **(20M)**

(a) Explain how performance of classifiers are measured?
(b) Define Gini Index and Discuss its purpose in decision tree algorithm.
(c) Discuss Naive Bayes algorithm with proper example.
(d) Enlist the steps in K- means clustering.
(e) Where can be Web Mining used? Justify your answer.
(f) How is Linear regression useful in Data Mining?

Q. 4 Attempt ANY FIVE from the following: **(15M)**

(a) Explain metadata repository in Data Warehouse.
(b) What do you mean by granularity of facts?
(c) Explain the term support and confidence with proper example.
(d) Elaborate the concept of decision tree.
(e) What does confusion matrix mean? Explain.
(f) How is data from world wide web mined?

****