Candidates are required to give their answers in their own words as far as practicable.
The figures in the margin indicate full marks.

System Programming

[Theory]

Group - A

Answer *any two* of the following: 2×15=30

1. a) What is the difference between syntax and semantic error?

   b) Explain the two-pass assembler in detail with block diagram.

   c) Explain details of the Lexical analysis phase of a compiler with example.

   3+7+5

2. a) What do you mean by ambiguity of grammar? Explain with example.

b) Generate any three intermediate representation for the following expression:

A=(B/C)*[(C+D)*B]*(E-C-100)

c) Explain absolute loader scheme with advantages and disadvantages.

5+4+6

3. a) What do you mean by Dynamic Loading?

b) Consider the context –free grammar:

S ->aX

X ->bX/bY

Y -> c

The symbols S, X, Y are non-terminals and S is the start symbol whiles a, b and c are terminal symbols.

i) Give the canonical collection of LR (0) items for the programme.

ii) Is this grammar SLR? Prove by constructing an SLR parsing table.

c) What are the differences between compiler and Interpreter?

d) Write a short note on scanning and parsing.   2+7+3+3

4. a) Write the function of the Symbol table of a compiler.

b) Explain how a Macro is different from Subroutine.

c) What is LEX used for? What is the difference between LEX and YACC?

d) Write a short note on Overlay.   3+3+5+4

GROUP B

Answer *any one* of the following: 1×10=10

5. Discuss the compilation process with a suitable example.   10

6. Explain the steps involved in the design of direct linking loader. 10

[Practical]

Answer *any one* of the following: 1×20=20

7. Write LEX program to Check whether the input is digit or not.

8. Write a LEX program to identify and count positive and negative number.

9. Write a YACC program to implement a calculator and recognize a valid arithmetic expression.

———————

Data Mining

[Theory]

Group - A

Answer *any two* of the following: 2×15=30

1. Briefly discuss about different descriptive and predictive data mining techniques. Write down the basic differences between these two techniques.

(10+5)

2. What do you mean by supervised learning? How supervised learning can be broadly categorized? Explain about these categories with proper examples. Briefly explain about K Nearest Neighbour algorithm with proper example.

(2+2+6+5)

3. How unsupervised learning is different from supervised learning? Briefly discuss about clustering and association with suitable examples. Give a diagrammatical representation of the steps involved in the knowledge discovery from data. (3+8+4)

4. Briefly discuss about different data pre-processing methods in data mining. Write a short note on association rule mining. State the Apriori property and state the two major drawbacks of the Apriori method. (6+6+3)

GROUP B

Answer *any one* of the following: 1×10=10

5. Briefly discuss about Support Vector Machine with suitable example. What do you mean by linear regression and logistic regression? Write down the applications of clustering. (5+3+2)

6. What do you mean by scalability issue in data mining? What are the approaches to minimize the scalability issue? Briefly explain measures of interestingness in data mining.

(2+3+5)

[Practical]

Answer *any one* of the following: 1×20=20

Transaction ID	Items Bought
1	Tea, Coffee, Cold Drink, Milk
2	Tea, Coffee, Cold Drink, Cake
3	Tea, Cold Drink, Eggs
4	Cake, Milk, Eggs, Tea
5	Tea, Coffee, Cold Drink, Milk, Cake, Eggs

7. From the above dataset, calculate support that "*if a person buys Tea, also buys Cake*". Also calculate the confidence that "*if a person buys Milk, also buys Tea*". Write down proper formulas of support and confidence before calculating.

20

7. The following table consists of training data. Construct a decision tree based on these data. Classify the records by the "Status" attribute. Write down the rules that can be generated from the obtained decision tree. 20

Department	Age range	Salary class	Status
Sales	Middle-aged	High	Senior
Sales	Young	Low	Junior
Sales	Middle-aged	Low	Junior
Systems	Young	High	Junior
Systems	Middle-aged	High	Senior
Systems	Young	High	Junior
Systems	Senior	High	Senior
Marketing	Middle-aged	High	Senior
Marketing	Middle-aged	Average	Junior
Secretary	Senior	Average	Senior
Secretary	Young	Low	Junior

8. Use the K-means algorithm and Euclidean distance to cluster the following data into 3 clusters:

X1(2,8), X2(2,5), X3(8,4), X4(10,4),

Y1(5,8), Y2(7,5), Y3(6,4),

Z1(1,2), Z2(4,9), Z3(5,10).

Suppose that the initial seeds (centre of each cluster) are X1, X4, Z2.

Run the K-means algorithm for 3 iterations. At the end of each iteration, show:

   i) The new clusters (i.e., the examples belonging to each cluster)

   ii) The centre of the new clusters

   iii) Draw a 10 by 10 space with all the 10 points and show the clusters after each iteration.

———