

## JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

(T-2 Examination April - 2019)

B. Tech. 8<sup>TH</sup> Semester

COURSE CODE: 11B1WCI832

MAX. MARKS: 25

COURSE NAME: INFORMATION RETERIVAL  
AND DATA MINING

COURSE CREDITS: 3

MAX. TIME: 1.5 Hrs

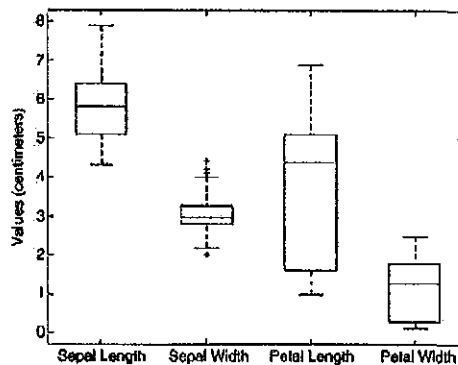
*Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.*

- Q.1 [CO 2] Explain the following terminology in relation to performance evaluation of classification algorithm with example: (5)
- Confusion Matrix
  - TP, FN, FP, TN
  - Accuracy
  - Precision
  - Recall
- Q.2 [CO 2] Why naïve Bayesian classification is called naïve? Briefly outline the major ideas and working of naïve Bayes classification with all its statistic formulations. (5)
- Q.3 [CO 2] Explain Entropy and Gini in terms of decision tree classification. Also provide comparison between these two using a diagram. (5)
- Q.4 [CO 2] Generate Association Rule for Transactional data illustrated below by applying Apriori Algorithm? (5)

Transactional Data for an AllElectronics Branch

TID	List of Item IDs
T100	11, 12, 15
T200	12, 14
T300	12, 13
T400	11, 12, 14
T500	11, 13
T600	12, 13
T700	11, 13
T800	11, 12, 13, 15
T900	11, 12, 13

- Q.5 [CO 2, 3] (a) Describe how a box plot can give information about whether the value of an attribute is symmetrically distributed. What can you say about the symmetry of the distributions of the attributes shown in illustration? (3+2)



- (a) Write an algorithm for k-nearest neighbour classification given k and n, the number of attributes describing each tuple.