1. Define the term “KNN” classification. Write two limitations of this classification.
2. Define the term True Positive and False Negative.
3. Define the terms: accuracy, recall, F-measures and precision.
4. Define tree based and rule based classification.
5. Write the process of classification of any one.
6. How ANN classifier works?
7. How Bayes classifier works?
8. How do you perform KNN? Write the limitations of KNN.
9. How do you validate a classification model?
10. What are the functions of ROC curve in validation?
11. How SVM classifier works?
12. Math on Draw dendrogram for hierarchical clustering.
13. If you have a data set with class attribute and a new data without class attribute. You want to predict the value of class attribute of new data. Write the process of classifying this new data using decision tree classification (Hunt’s Algorithm).
14. If you have a data set with class attribute and a new data without class attribute. Write the process of classifying this new data using ensemble method.
15. If you have a data set with class attribute and a new data without class attribute. Write the process of classifying this new data using KNN.
16. Math on validation.
17. In constructing a decision tree, how do we select an attribute and when do we stop the further expansion of the tree?
18. On what principal Bayesian classifier has been built?
19. Math on Gini Index
20. What are the advantages of tree based classification?
21. What are the main principal of Bayesian Classification?
22. What are the sequential steps in doing classification of a set of data?
23. What are the use of ROC curve?
24. What are the uses of support vector machine (SVM)?
25. What do you mean by the term precision and recall? When do we use these?
26. What do you mean by rule base classification?
27. What is decision tree classification?
28. What is ensemble method of classification? Explain with pictorial example.
29. What is the main principal of Gini index? – explain. When we have to use Gini index in splitting?
30. When we have to use Gini index in splitting?
31. Which classification technique will you use?
32. Why and when do researchers like to use SVM classifier?
33. Why and when do we use Gini coefficient or entropy?
34. Why entropy is used instead of Gini index?
35. Write the algorithm of K-nearest neighbour classification
36. Write the limitations of KNN.
37. What is K-nearest neighbour classification and k-means clustering?
38. Define hierarchical clustering with example.
39. Describe the steps of K-means clustering.

40. What are the different types of clustering?
41. What are the processes of density based clustering?
42. What do you mean by centeroid in k-means clustering
43. What do you mean by clustering? What are the different types of clustering?
44. What is clustering? Describe the steps of K-means clustering.
45. What is the procedure of validating k-means clustering?
46. Write the steps of k-means clustering?
47. Write the steps of k-means clustering?
48. Write two limitations of k-means clustering. How can you minimize these limitations?
49. Explain with a pictorial example of Core Point, Noise Point and Border Point.
50. In density based clustering, how do you select epsilon and distance? What are the logic of this process?
51. What is the basic principal of DBSCAN clustering?
52. What is the process of validating a density based clustering?
53. When we need to use density based clustering?
54. When we need to use density based clustering?
55. Write a situation where DBSCAN clustering is appropriate.
56. Write one application of DBSCAN clustering.
57. Data may affected by various kind of reasons. These reasons we may define as data quality problems. Answer the following questions: Explain these reasons with small examples.
58. Explain different types of data that we face in data mining.
59. Explain, how do you discretize a numeric attribute? i.e. Income
60. How can we detect problems with the data?
61. How do you discretize a numeric attribute? i.e. Age
62. If you have missing data and noise exist in your data then what are the steps you should take?
63. List different types of attributes with their general properties.
64. To calculate dissimilarity between two data objects you can use Euclidian Distance and Mahalanobis Distance. Which one will you prefer and why?
65. To calculate similarity or dissimilarity between two data objects which formulas you can use? Explain their differentials.
66. What are the different methods of calculating similarity and dissimilarity?
67. What are the different types of data set available? Give an example of each type.
68. How do you calculate distance between two clusters?
69. How hierarchical clustering helps to construct other clustering techniques?
70. Write procedure of hierarchical clustering with a data example.
71. Write some applications of hierarchical clustering?
72. Write the algorithm of hierarchical clustering.
73. Write the process of hierarchical clustering in your own words.
74. Give an example of data where Data Mining techniques need to apply to extract hidden and unknown information.
75. Define Data Mining? There are two types of Data mining techniques: Predictive and descriptive data mining- give example of these two.
76. Define the term data mining. Give an example of predictive data mining.

77. Non-trivial extraction of implicit, previously unknown and potentially useful information from data is called Data Mining. There are several tasks that we employ for mining; both classification and clustering. Answer the following questions: Give some examples of data where Data Mining techniques need to apply to extract hidden and unknown information.
78. What are the difference between classification and clustering?
79. What do you mean by supervised and unsupervised classification?
80. What is data mining and why is it an important discipline?
81. Why do we divide data in two parts before data mining starts?
82. Write the list of predictive data mining. How anomaly detection is one kind of data mining?
83. What is Data Mining? Why is data mining important in our daily life?
84. Before applying data mining techniques, data processing techniques need to apply. Explain some data processing techniques.
85. Explain different distance measures.
86. What is OLAP? Why do we need OLAP? Define the term “Slicing” and “Dicing”.
87. When do we need to use discretization and binarization?
88. When will you use Jaccard Coefficient and Cossine Similarity Index?
89. Why do we apply aggregation on data?