Francis Xavier Engineering College, Tirunelveli – 627 003
(An Autonomous Institution)
Department of Computer Science and Engineering
Year/ Semester: IV/VII **Academic Year: 2023-2024/ODD**
Course Code/Title: 19CS7701 DATA MINING– Set 1
CAT 2 – Question Bank

Q.No.	Question
	Unit 3
1.	Define association and correlations.
2.	Define an efficient procedure for cleaning the noisy data.
3.	Define data mining query language.
4.	What is the purpose of feature selection?
	Unit 4
1.	List the major clustering methods.
2.	What are ways of reducing dimensionality?
3.	How does decision tree induction algorithm work?
4.	Illustrate support vector machine with example.
5.	Define Lazy learners with an example.
6.	List the major classification methods.
7.	List the major clustering methods.
	Unit 5
1.	What are the advantages and disadvantages of logistic regression?
2.	How does temporal database differ from regular database?
3.	What are the advantages of spatial database?
4.	What are the applications of data mining?
5.	What is the difference between time series and sequential data?
6.	Differentiate data mining and Statistics

PART – B (5 x 13 = 65 Marks)

Q.No.	Question
	Unit 3
1.	What is data Pre-processing? Explain the various data pre-processing techniques.
2.	Generalize and discuss about association rule mining with examples and state how association mining to correlation analysis is dealt with.
3.	Demonstrate in detail about data mining steps in the process of knowledge discovery.
	Unit 4
1.	Describe in detail about the following Classification methods. ● Bayesian classification ● Classification by Back propagation.
2.	Define classification? With an example explain how support vector machines can be used for classification.

3.	Develop an algorithm for classification using decision trees. Illustrate the algorithm with a relevant example.
4.	Explain the following: ● Hierarchical based method. ● Density based methods.
5.	What is outlier mining important? Briefly describe the different approaches behind statistical–based outlier detection, distance-based outlier detection and deviation-based outlier detection.
6.	What is grid-based clustering? With an example explain an algorithm for grid-based clustering.
	Unit 5
1.	What is temporal database? Explain temporal database with example.
2.	What is web mining? Explain the various types of web mining methods.
3.	What is Multimedia database? Explain about multimedia database with example.
4.	Explain about text mining and discuss about the challenges in text mining.
5.	Discuss about spatial databases with example.
6.	What is logistic regression? Discuss in detail about various types of logistic regression.

PART – C (1 x 15 = 15 Marks)

Q.No.	Question
1.	Discuss the steps in K-means algorithm and evaluate the following table using K-means.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

2.	Analyse and elaborate the current trends in data mining in any three fields. 1. Financial data analysis 2. Biological data analysis 3. Telecommunication industry 4. Intrusion detection

	5. Retail industry
3.	Explain and Apply the Apriori algorithm for discovering frequent item sets of the table. Trans ID | Items Purchased T1 | I1, I2, I5 T2 | I2, I4 T3 | I2, I3 T4 | I1, I2, I4 T5 | I1, I3 T6 | I2, I3 T7 | I1, I3 T8 | I1, I2, I3, I5 T9 | I1, I2, I3 minimum support count is 2 and minimum confidence is 60%. Illustrate each step of the Apriori Algorithm.