Module 1: Question Bank

Notes:

1. **Each question carries 2 marks each in Part A**
2. **Each question carries 8 marks each in Part B**

PART A

Q#	Question	Marks
1	What is data mining? Differentiate between the data mining and data retrieval.	2
2	List and explain the data mining motivational challenges.	2
3	Identify the different types of attributes for the dataset given in table1.	2

Table 1. Borrower Dataset

TId	Refund	Marital Status	Taxable Income	Default Borrower
1	Yes	Single	125k	No
2	No	Married	100k	No
3	No	Single	70k	No
4	Yes	Married	120k	Yes
5	Yes	Divorced	220k	No

Q#	Question	Marks
4	What are the general characteristics of a dataset?	2
5	In brief describe the data quality issues.	2
6	For the given data, find out the Pearson's correlation coefficient. X=(-3, 6, 0, 3, -6) Y=(1, -2, 0, -1, 2)	2
7	For the given data, find out the Euclidean distance between the points.	2

Table 2. x-y coordinates of the given points.

Point	x-Coordinate	y-Coordinate
P1	0	2
P2	2	0
P3	3	1
P4	5	1

Q#	Question	Marks
8	What is a metric? Describe.	2
9	In brief describe the four major data mining task.	2
10	Briefly explain the issues in calculating the proximity.	2

PART B

			Marks
1	a.	Describe KDD process in brief.	4
	b.	Describe the different types of data.	4
2	a.	Discuss the different data pre-processing techniques. For the data given below, in table 3, find out the 1. Simple matching coefficient 2. Jaccard Co-efficient 3. Cosine Similarity	4
	b.	Table 3. Document Vectors.	4

Vectors	A	1	0	0	0	0	0	0	0	0	0
	B	0	0	0	0	0	0	1	0	0	1