Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	Which of these is correct about data mining?	It involves processes like Data Transformation, Data	It is a procedure in which knowledge is mined from data.	It is a procedure using which one can extract information	All of the above	D	CO1	L3
A	obj	1	_____ are the data objects that don't comply with the general model or behaviour of the available data:	Evolution Analysis	Outlier Analysis	Classification	Prediction	B	CO1	L3
A	obj	1	The issues of "Scalability and efficiency of the data mining algorithms" come under	User Interaction and Mining Methodology Issues	Diverse Data Types Issues	Performance Issues	None of the above	C	CO1	L3
A	obj	1	The primary use of data cleaning is	Removing the noisy data	Correction of the data inconsistencies	Transformations for correcting the wrong data	All of the above	D	CO1	L3
A	obj	1	The classification of the Data Mining System consists of	Machine Learning	Information Science	Database Technology	All of the above	D	CO1	L3
A	obj	1	Out of the following, which one is the proper application of data mining	Fraud Detection	Market Management and Analysis	Risk Management & Corporate Analysis	All of the above	D	CO1	L3
A	obj	1	The class under study in Data Characterization is known as	Initial Class	Target Class	Final Class	Study Class	B	CO1	L3
A	obj	1	The issue of "Handling complex and relational types of data" comes under:	Performance Issues	User Interaction and Mining Methodology Issues	Diverse Data Types Issues	None of the above	C	CO1	L3
A	obj	1	The initial steps concerned in the process of knowledge discovery is	Data Selection	Data Integration	Data Cleaning	Data Transformation	C	CO1	L3
A	obj	1	Which of the following is generally used in finding hidden structure and patterns in a given unlabelled data	Supervised learning	Unsupervised learning	Reinforcement learning	None of the above	B	CO1	L3
A	obj	1	KDD stands for	Knowledge Discovery Database	Knowledge Definition Data	Knowledge Data Discovery	Knowledge Data Definition	A	CO1	L3
A	obj	1	Which of the following statements is true about the classification?	It is a measure of accuracy	It is a subdivision of a set	It is the task of assigning a classification	None of the above	B	CO1	L3
A	obj	1	Which of the following process uses intelligent methods to extract data patterns?	Data mining	Text mining	Warehousing	Data selection	A	CO1	L3
A	obj	1	You are given a seismic activity of a country and your want to predict the magnitude of an earthquake , this is an exampke of	Supervised learning	Unsupervised Learning	Dimension reduction	None of the above	A	CO1	L3
A	obj	1	You want to predict number of newborns depending upon the size of the population, this is an example of	Classification	Regression	Clustering	SEM	B	CO1	L3
A	obj	1	Examples of Nominal can be	ID Numbers, eye color, zip codes	Rankings, taste of potato chips, grades, height	Calendar dates, temperatures in celsius or Fahrenheit,	The temperature in Kelvin, length, time, counts	A	CO1	L3
A	obj	1	Examples of Interval can be	ID Numbers, eye color, zip codes	Rankings, taste of potato chips, grades, height	Calendar dates, temperatures in celsius or Fahrenheit,	The temperature in Kelvin, length, time, counts	D	CO1	L3
A	obj	1	What are some examples of data quality problems	Noise and outliers	Duplicate data	Missing values	All of the Above	D	CO1	L3

Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	Under fitting happens due to	A fewer number of features	Data has a high variance	No use of regularization	All of the above	A	CO1	L3
A	obj	1	Which of the following is true about outliers	Data points that deviate a lot from normal observations	Can reduce the accuracy of the model	Both A and B	None of the above	C	CO1	L3
A	obj	1	Which algorithm does not require feature scaling	Naive Bayes	Decision Tree	Both A and B	None of the above	D	CO2	L3
A	obj	1	Which of the following is defined as the Euclidean distance measure	Finding the solution for a problem simply by summarising all	A KDD process stage in which new data is added to the existing selection	Both A and B	None of the above	A	CO2	L3
A	obj	1	Is discrimination between spam and ham emails a classification task	Yes	No			A	CO2	L3
A	obj	1	From where are classification rules extracted	Branches	Siblings	Decision Tree	Root Node	C	CO2	L3
A	obj	1	What is classification in data mining?	Setting up a target data	Data mining procedure to sort data	A method to find data	Generalizing structures	D	CO2	L3
A	obj	1	What is Summarization in data mining?	Setting up a target data	Data mining procedure to sort data	A method to find data	To represent the derivate data with visualization and reports.	D	CO2	L3
A	obj	1	What is regression	Database technologies	Information used to calculate funds	The statistical method used in finance, investing,	Mathematical method to derive cost	C	CO2	L3
A	obj	1	How to use regression analysis in data mining?	Database technology	It is used to calculate funds	It is used to derive cost	It is used to predict a range of numeric values	D	CO2	L3
A	obj	1	What is a Dummy Variable Trap?	Multicollinearity among the dummy variables	One variable predicts the value of other	Both A and B	None of the above	C	CO2	L3
A	obj	1	How to handle the missing values in the dataset?	Dropping the missing rows or columns	Imputation with mean/median/mode value	Taking missing values into a new row or column	All of the above	B	CO2	L3
A	obj	1	Identify the distance matrix used in KNN Algorithm_____	Euclidean Distance.	Cosine Similarity	Both A and B	None of the above	A	CO2	L3
A	obj	1	_____ algorithm is widely used for different kinds of learnings because of its uncomplicated and easy to apply nature	KNN	Clustering			A	CO2	L3
A	obj	1	_____in KNN is a parameter that refers to the number of nearest neighbours to include in the majority of the voting process.	K	First N	Last N	None of the above	A	CO2	L3
A	obj	1	Is cross validation another way to choose k	Yes	No			A	CO2	L3
A	obj	1	____ is the mathematical likelihood that something will occur	Classification	Probability	Naive Bayes Classifier	None of the above	C	CO2	L3
A	obj	1	The Naive Bayes Classifier is a ____ in probability	Technique	Process	Classification	None of the above	A	CO2	L3

Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/ B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.	TRUE	FALSE			A	CO2	L3
A	obj	1	Which of the following statements about the K-means algorithm are correct?	The K-means algorithm is sensitive to outliers	The centroids in the K-means algorithm may not be any observed data points	Both A and B	None of the above	C	CO2	L3
A	obj	1	Decision trees can handle	High dimensional data	low diamesional data	medium dimensional data	None of the above	A	CO2	L3
A	obj	1	Decision-tree algorithm falls under the category of	Supervised learning	Unsupervised learning	Reinforcement learning	None of the above	A	CO2	L3