SNS College of Technology, Coimbatore-35.
(Autonomous)
Academic Year 2023-2024(ODD)
Fifth Semester
Computer Science and Engineering
19CSE301 – INTRODUCTION TO DATA SCIENCE
Part-A (2Marks)

Five Units

1	Define Generalizing from Data
2	What are the types of data generalization?
3	Define Automated generalization
4	Define Declarative Generalization.
5	What is Rectangular data?
6	Define relational database.
7	Write create command for table and explain SQL.
8	Define Indexes.
9	Define slicing.
10	Define sorting
11	What is Data Representation?
12	List out the Data Representation.
13	What is data Quality?
14	Write three measures of Data Quality.
15	What is exploratory Data Analysis?
16	Write about Univariate Data Analysis
17	What are the steps involved in exploratory data Analysis.
18	Write down the area of text mining.
19	What is Text Analytics?
20	List out the techniques of Text Analytics
21	Define Regular Expression.
22	What is CSS?
23	Define REST and abbreviate.
24	What is client server process?
25	List out the expression used in Xpath.

26	List out the important features in Xpath,
27	Define Naïve Bayes Classifier.
28	Write the Bayes rule or formula.
29	Write down the advantage and disadvantage in Bayes Theorem
30	List out the application in Bayes Theorem.
31	What is Regression in probabilities.
32	List out the types of regression
33	What is the Logistic Regression?
34	List out the algorithms used in logistic regression
35	Write down the assumption in linear regression.
36	Define Multinominal regression.
37	List out the types in multinominal regression.
38	Define polynomial regression.
39	Write the equation for polynomial regression.
40	List out the steps in polynomial regression
41	List out the types of polynomial regression.
42	Draw the sigmoid function with values.
43	Define LDA
44	Write formula and explain the simple linear regression model
45	List out the variable used in the regression.
46	Define P hacking
47	Define False discovery rate.
48	Define Multiple testing Problem.
49	Define dimensionality reduction.
50	List out the Components of dimensionality reduction.
51	List out the ways used for problem.
52	List out the methods in Dimensionality reduction.
53	Define PCA.
54	List out the application in PCA/SVD.
55	List out the advantage in dimensionality reduction.
56	List out the disadvantage in dimensionality reduction.
57	Define decision tree.
58	Define CART algorithm.
59	List out then types of Decision Tree.
60	List out the terminologies used in Decision tree
61	Define ASM.
62	Define Entropy and its formula

63	Define Gini Index and its formula.
64	List out the techniques in ASM.

Part B

Five Units

1	Explain briefly about Indexes and types with using commands.
2	Explain slicing and sorting using data science.
3	Apply and plot the information using data science
4	Summarize the data representation using data science technologies.
5	Explain brief about the quality of data.
6	Explain the types of Exploratory data analysis.
7	Illustrate the steps involved in Exploratory Data Analysis
8	Explain brief about mining and process the text using data science.
9	Illustrate the steps and techniques used in text analytics.
10	Explain about client server communication takes place in web technologies.
11	Summarize the REST used in data science
12	Compare the XQuery and XSLT using expression in data science.
13	Illustrate the Naïve bayes Theorem used in expressions.
14	Summarize the file handling in single computer with neat diagram.
15	Explain brief about the regression on probabilities
16	Explain the classification and types used in logistic regression.
17	Explain the classification used in logistic Model.
18	Construct the loss function and cost function for logistic model.
19	Evaluate and fitting the logistic model.
20	Explain brief about multiclass classification used neat diagram.
21	Construct the diagram to visualize the data to the end users.
22	Explain briefly about P Hacking.
23	Illustrate the dimensionality reduction in the part of replicability.
24	Explain brief about PCA and using singular value
25	Construct the decision tree using some example
26	Explain brief about the Random forest.
27	Discuss in detail about data science Life Cycle
28	In a school 600 students are studying various types of language as English, Tamil, Malayalam list the count of the students studying various languages with represent the graph using data science.
29	If you are one of those who is having an interest in natural language processing then this use case is for you. The idea is to train a machine-learning model to generate emojis based on the input text. Then this machine learning model can be used in training Artificial Intelligent Chatbots.
30	Evaluate whether the weather is sunny, then the Player should play or not.

			Outlook	**Play**
		0	Rainy	Yes
		1	Sunny	Yes
		2	Overcast	Yes
		3	Overcast	Yes
		4	Sunny	No
		5	Rainy	Yes
		6	Sunny	Yes
		7	Overcast	Yes
		8	Rainy	No
		9	Sunny	No
		10	Sunny	Yes
		11	Rainy	No
		12	Overcast	Yes
		13	Overcast	Yes
31	Develop the data to visualize the chart you may have a 3-class classification problem of a set of fruits to classify as oranges, apples, or pears with a total of 100 instances. A total of 80 instances are labeled with Class-1 (Oranges), 10 instances with Class-2 (Apples), and the remaining 10 instances are labeled with Class-3 (Pears).			
32	Construct the possible solution using a decision tree. 1. If the travel cost per km is expensive, the person uses a car. 2. If the travel cost per km is the standard price, the person uses a train. 3. If the travel cost is cheap, the decision tree needs to ask the next question about the gender of the person. If the person is a male, then he uses a bus. 4. If the gender is female, the decision tree needs to ask again how many cars she owns in her household. 5. If she has no car, she uses a bus, Otherwise, she uses the train.			