NAVRACHANA UNIVERSITY
a UGC recognized University

School:	School of Engineering and Technology
Program/s:	B.Sc(Data Science)
Year: 2nd	**Semester:** 4th
Examination:	End Semester Examination
Examination year:	May – 2023

Course Code: DS210   **Course Name:** Machine Learning
Date: 16/05/2023
Time: 10:00 am to 12:00 pm

Total Marks: 40
Total Pages: 07

Instructions:

a) Attempt All the Questions
b) No Calculator is allowed

Q. No.	Details	Marks	CO	BTLO
	Attempt All the Questions(1 marks each)	40		
Q.1	Machine Learning model should be fast, accurate and Scalable, A. True B. False	1	CO2	BTLO2
Q.2	The model learns and updates itself through reward/punishment in case of A. Supervised learning algorithm B. Unsupervised learning algorithm C. Semi-supervised learning algorithm D. Reinforcement learning algorithm	1	CO3	BTLO3
Q.3	Which of the following are not classification problems? (Select two) * (a) Predicting price of house (b) Predicting patient has tumor (c) Predicting who will hold the title in football league (d) Predicting percentage of student for next semester	1	CO4	BTLO4
Q.4	Some test results are normally distributed, with a mean of 18 and a standard deviation of 6. What percentage of test takers scored between 18 and 24? (a) 20% (b) 22% (c) 34% (d) None of the above	1	CO5	BTLO5
Q.5	The association between the number of years an employee has been with a firm and the person's pay is 0.75. What can be stated regarding employee pay and years of service? (a) There is no relationship between salary and years worked (b) The majority of employees have been with the company a long time (c) Individuals that have worked for the company the longest have lower salaries (d) Individuals that have worked for the company the longest have higher salaries	1	CO2	BTLO1
Q.6	Computers are excel at learning _____ *	1	CO3	BTLO2

	(a) facts (b) concepts (c) procedures (d) principles			
Q.7	Which of the following are unsupervised problem? (Select two) * (a) Google news (b) Rain predictor (c) Customer segmentation (d) Tumor prediction	1	CO2	BTL03
Q.8	This is the cleaning/transforming the data set in the supervised learning model (a) Problem Identification (b) Identification of Required Data (c) Data Pre-processing (d) Definition of Training Data Set	1	CO2	BTL04
Q.9	A multiple regression model has: * (a) Only one independent variable (b) More than one independent variable (c) More than one dependent variable (d) None of the above	1	CO1	BTL05
Q.10	Real-Time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are applications of .... * (a) Reinforcement Learning (b) Supervised Learning: Classification (c) Unsupervised Learning: Regression (d) None of the above	1	CO1	BTL05
Q.11	Structured representation of raw input data to meaningful ___ is called a model. (a) pattern (b) data (c) object (d) none of the above	1	CO4	BTL05
Q.12	The process of supervised machine learning is as good as the data	1	CO4	BTL03

	used to train it A. False B. True			
Q.13	Machine Learning is a field of AI consisting of learning algorithms that _____ (a) At executing some task (b) Over time with experience (c) Improve their performance (d) All of the above	1	CO3	BTLO3
Q.14	What are successful applications of Machine Learning? (a) Learning to recognize spoken words (b) Learning to drive an autonomous vehicle (c) Learning to classify new astronomical structures (d) All of the above	1	CO3	BTLO3
Q.15	Logistic regression is a ……….. regression technique that is used to model data having a ……….. outcome. (a) linear, numeric (b) linear, binary (c) nonlinear, numeric (d) nonlinear, binary	1	CO2	BTLO2
Q.16	Which of the following statement is true about prediction problems? (a) The output attribute must be numeric. (b) he output attribute must be categorical (c) The resultant model is designed to determine future outcomes (d) The resultant model is designed to classify current behavior	1	CO2	BTLO2
Q.17	If machine learning model output involves target variable, then that model is called as predictive model. A. True B. False	1	CO1	BTLO1
Q.18	The probabilistic approach used in machine learning is closely related to: (a) Statistics (b) Physics (c) Mathematics (d) Psychology	1	CO1	BTLO1

Q.19	Common classes of problems in machine learning is .... *	1	CO1	BT
	(a) Clustering			
	(b) Regression			
	(c) Classification			
	(d) All of the above			
Q.20	An alternative name for an output attribute	1	CO1	BTL06
	(a) predictive variable			
	(b) independent variable			
	(c) estimated variable			
	(d) dependent variable			
Q.21	Let us consider two examples, say 'predicting whether a tumour is malignant or benign' and 'price prediction in the domain of real estate'. These two problems are same in nature A. True B. False	1	CO1	06
Q.22	Let's say your model is overfitting. Which of the following is NOT a suitable method for attempting to decrease overfitting?	1	CO1	BTL06
	(a) Increase the amount of training data			
	(b) Improve the optimization algorithm being used for error minimization.			
	(c) Decrease the model complexity.			
	(d) Reduce the noise in the training data			
Q.23	Which data is used to tune the parameters of supervised learning Model	1	CO1	06
	(a) training			
	(b) test			
	(c) verification			
	(d) validation			
Q.24	What is Machine learning?	1	CO1	BTL01
	(a) The autonomous acquisition of knowledge through the use of computer programs (b) The autonomous acquisition of knowledge through the use of manual programs			

(c) The selective acquisition of knowledge through the use of computer 1programs

(d) The selective acquisition of knowledge through the use of manual programs

Q.25	Temperature is a _____ . (a) Interval data (b) Ratio data (c) Discrete data (d) None of the above	1	CO1	BTLO4
Q.26	For understanding relationship between two variables, ____ can be used. (a) Box plot (b) Scatter plot (c) Histogram (d) None of the above	1	CO1	BTLO4
Q.27	Out of 200 emails, a classification model correctly predicted 150 spam emails and 30 non-spam emails. What is the accuracy of the model? (a) 10% (b) 90% (c) 80% (d) none of the above	1	CO1	BTLO4
Q.28	This is the first step in the supervised learning model. (a) Problem Identification (b) Identification of Required Data (c) Data Pre-processing (d) Definition of Training Data Set	1	CO1	BTLO2
Q.29	Machine learning techniques differ from statistical techniques in that machine learning methods (a) typically assume an underlying distribution for the data (b) are better able to deal with missing and noisy data (c) are not able to explain their behavior. (d) have trouble with large-sized datasets	1	CO1	BTLO2

Q.30	PCA is a technique for	1	CO1	B
	(A) Feature extraction			
	(B) Feature construction			
	(C) Feature selection			
	(D) None of the above			
Q.31	Which of the following statements regarding outliers is correct? *	1	CO1	BTL01
	(a) Outliers should be identified and removed from a dataset			
	(b) Outliers should be part of the training dataset but should not be present in the test data			
	(c) Outliers should be part of the test dataset but should not be present in the training data			
	(d) The nature of the problem determines how outliers are used			
Q.32	Data can be broadly divided into following two types (Select Two) *	1	CO1	BTL01
	(a) Qualitative			
	(b) Speculative			
	(c) Quantitative			
	(d) None of the above			
Q.33	Two events A and B are called mutually exclusive if they can happen Together	1	CO1	BTL01
	A. True			
	B. False			
Q.34	A term used to describe the case when the independent variables in a multiple regression model are correlated is	1	CO1	01
	(a) regression			
	(b) correlation			
	(c) multicollinearity			
	(d) none of the above			
Q.35	Training data on which the algorithm run is called as _____.	1	CO1	BTL01
	(a) Program			
	(b) Model Training			
	(c) Training Information			

				1	CO1	BTLO1

(d) Learned Function

Supervised learning differs from unsupervised clustering. Supervised learning requires

    (a) At least one input attribute

    (b) Input attributes to be categorical

    (c) At least one output attribute

    (d) Output attributes to be categorical

Q.37 Machine learning is a _____ field. *   1   CO1   BTLO4

    (a) Inter-disciplinary

    (b) Single

    (c) Multi-disciplinary

    (d) All of the above

Q.38 What is the most common issue when using Machine Learning?   1   CO1   BTLO4

    (a) Poor Data Quality

    (b) Lack of skilled resources

    (c) Inadequate Infrastructure

    (d) None of the above

Q.39 Which of the following is not TRUE for regression? *   1   CO1   BTLO3

    (a) It relates inputs to outputs

    (b) It is used for prediction

    (c) It may be used for interpretation

    (d) It discovers causal relationships

40 .... is not a machine learning algorithm.   1   CO1   BTLO3

    (a) SVG

    (b) SVM

    (c) Random forest

    (d) All of the above

************End of Question Paper************