



BRAINWARE UNIVERSITY

Term End Examination 2023-2024 Programme – MCA-2022 Course Name - Applied Data Science **Course Code - MCA401B** (Semester IV)

Time: 2:30 Hours Full Marks: 60

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1.	Choose the correct alternative from the following	ng:
(i)	Predictive analysis is defined as branch of	analysis.
(ii)	a) Advanced c) Both of these	b) Core d) None of these alysis process contains.
/:::\	a) 5 c) 7 is identified to make predictions	b) 6 d) 8 about unknown future events.
(iii)	a) Predictive analysis	b) Descriptive analysis
(iv)	c) Both of these Bayesian classifiers is defined as	d) None of these
	a) A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.c) An approach to the design of learning	b) Any mechanism employed by a learning system to constrain the search space of a hypothesisd)
	algorithms that is inspired by the fact that	Y

b) John Chambersii

None of these

- d) None of these
- b) The process of finding a solution for a problem simply by enumerating all possible

(vi) Euclidean distance measure is calculated in

a) A stage of the KDD process in which new data is added to the existing selection

when people encounter new situations,

they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation

(v) Who developed R? a) Dennis Ritchie

c) Biarne Stroustrup

- c) The distance between two points as calculated using the Pythagoras theorem
- (vii) Hidden knowledge relates to
 - a) A set of databases from different vendors, possibly using different database paradigms
 - c) Information that is hidden in a database and that cannot be recovered by a simple SQL query
- (viii) Enumeration is related to
 - a) A stage of the KDD process in which new data is added to the existing selection
 - c) The distance between two points as calculated using the Pythagoras theorem
- (ix) KDD is interpreted as (Knowledge Discovery in Databases)
 - a) Non-trivial extraction of implicit previously unknown and potentially useful information from data
 - c) collection of interesting and useful patterns in a database
- (x) What is Naive prediction?
 - a) A class of learning algorithms that try to derive a Prolog program from example
 - c) A prediction made using an extremely simple method, such as always predicting the same outpu
- (xi) Determine the wrong statement
 - a) The trapezoidal rule is used to compute the area under the ROC curve
- c) An argument, para, is used to pick the model fitting technique (xii) To evaluate the minimum or the maximum of a function, we set the gradient to zero
- because a) The value of the gradient at extrema of a
 - function is always zero c) Both The value of the gradient at extrema of a function is always zero and Depends on the type of problem
- (xiii) What is the purpose of performing cross-validation?
 - a) To assess the predictive performance of the models
 - c) oth To assess the predictive performance of the models and To judge how the trained model performs outside the sample on test

- order and then testing them
- d) None of these
- b) An approach to a problem that is not guaranteed to work but performs well in most cases

solutions according to some pre-defined

- d) None of these
- b) The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them
- None of these
- b) Set of columns in a database table that can be used to identify each record within this table uniquely
- d) None of these
- b) A table with n independent attributes can be seen as an n-dimensional space
- d) None of these
- b) For regression, the relationship between each predictor and the outcome is evaluated
- d) All of these
- b) Depends on the type of problem
- None of these
- b) To judge how the trained model performs outside the sample on test data
- None of these

d)

- (xiv) Which of the following is distinguished as example of feature extraction?
 - a) Constructing bag of words vector from an emai
- b) Applying PCA projects to a large highdimensional dat

 c) Removing stopwords in a sentence (xv) In which agent the problem generator is judged a) Learning agent c) Reflex agent 	d) All of theseged ?b) Observing agentd) None of these		
Group-B			
(Short Answer Type Questions) 3 x 5			
 Compare supervised and unsupervised learning. How can outlier values can be determined? What do you conclude about true positive rate and false-positive rate? Define bias-variance trade-off. Develop life cycle of Data Science. OR Express the use of Data science with example.			
Group-C			
(Long Answe	5 x 6=30		
 Construct an example with an alternative hyp Differentiate between Data Analytics and Data Justify the concept of central limit theorem. Estimate noises in terms of data science. Compare and explain different types of Mean Discuss left-skewed and right-skewed distribut Discuss Descriptive and Inferential Statistics and 	with example. tions. OR	(5) (5) (5) (5) (5) (5)	