Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. -------- function is used to add a title to each axis instance in a figure.

A : set_title()

B : get_title()

C : set_label()

D : title()

Q.no 2. ---------- provides arange of supervised and un-supervised learning algorithms via consistant interface in python

A : Pandas

B : Numpy

C : Scikit-Learn

D : image

Q.no 3. The ---------- attribute specifies the number of dimensions or axes of the array.

A : ndarray.size

B : ndarray.dtype

C : ndarray.ndim

D : ndarray.axes

Q.no 4. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 5. ---------- submodule of scipy is dedicated to image processing.

A : ndarray

B : spatial

C : ndimage

D : special

Q.no 6. If number of input features are 3 then optimal hyperplane in support vector machine is ----------

A : Single point

B : Line

C : 2-D Plane

D : Non linear line

Q.no 7. ---------- is an example of human generated unstructured data.

A : Text files

B : Satellite data

C : Sensor data

D : Seismic imagery data

Q.no 8. ---------- must be installed before you use scikit learn

A : Matlab

B : Scilab

C : Scipy

D : Numpy

Q.no 9. The procedure to organize items of a given collection into groups based on some similar features called as ----------

A : Regression

B : Clustering

C : Ddecion Trees

D : Association

Q.no 10. In statistics, a population consists of ----------

A : All People living in a country.

B : All People living in the city.

C : All subjects or objects whose characteristics are being studied.

D : Part of whole dataset

Q.no 11. Which function is used to give title for the axes.

A : plt.title()

B : plt.xlabel()

C : plt.ylabel()

D : plt.xscale()

Q.no 12. ---------- function is used to plot a histogram using matplotlib library

A : hist()

B : bar()

C : pie()

D : scatter()

Q.no 13. Which of the following is measure used in decision trees while selecting splliting criteria that partitions data into the best possible manner.

A : Probability

B : Gini Index

C : Regression

D : Association

Q.no 14. Email data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 15. Which of the following is not a type of clustering algorithm?

A : Density clustering

B : K-Mean clustering

C : Centroid clustering

D : Simple clustering

Q.no 16. ------ answers the questions like " How can we make it happen?"

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 17. ---------- data does not fits into a data model due to variatins in contents.

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 18. ---------- function multiply two matrices in numpy.

A : prod()

B : mult()

C : dot()

D : *

Q.no 19. ---------- is a general purpose array-processing package provides a high performance multi-dimentional array object and tools for working with these arrays.

A : NumPy

B : SciPy

C : sklearn

D : None of these

Q.no 20. -------- library is built on the top of Numpy, SciPy and Matplotlib

A : Sympy

B : Scikit

C : Pandas

D : Numpy

Q.no 21. The last element of ndarray is indexed by ----------

A : 0

B : -1

C : 1

D : -2

Q.no 22. ----------the step is performed by data scientist after acquiring the data.

A : Data Cleansing

B : Data Integration

C : Data Replication

D : Data loading

Q.no 23. ---------- function is used to save an array as in image file.

A : matplotlib.pyplot.image()

B : matplotlib.pyplot.imread()

C : matplotlib.pyplot.imwrite()

D : matplotlib.pyplot.imsave()

Q.no 24. ---------- is unsupervised machine learning technique.

A : KNN

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 25. What is correct syntax to generate inetegers between 10 to 30

A : x=numpy.arange(10,30)

B : x=numpy.array(10,30)

C : x=numpy.arange(10,31)

D : x=arange(10,31)

Q.no 26. ---------- function used to get arrays elementwise remainder of division

A : numpy.divide(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.reminder(x1,x2)

Q.no 27. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 28. A ---------- is a supervised machine learning algorithm which relies on the assumptiion of feature independent to classify input data.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 29. What is the use of following function? Plt.xlabel("Total Marks")

A : Gives label to X-Axis

B : Gives label to Y-Axis

C : Gives title to figure

D : Add text to figure

Q.no 30. Pandas provide ---------- function as the entry point for all standard database join operations while merging two DataFrame objects.

A : concat()

B : replace()

C : merge()

D : add()

Q.no 31. Data generated on twitter is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 32. ---------- is an excellent 2D and 3D graphics library for generating scientific figures?

A : Pandas

B : Numpy

C : matplotlib

D : ndarray

Q.no 33. Support(B) =

A : (Transacions containing (B)) / (Total Transactions)

B : (Transacions containing (B)) / 100

C : (Total Transactions) / (Transacions containing (B))

D : 100/ (Transacions containing (B))

Q.no 34. ---------- is an example of semi structured data

A : NoSQL data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 35. ---------- is raster graphic format with lossless compression.

A : EPS

B : PDF

C : PNG

D : PS

Q.no 36. ----------is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 37. ---------- is a form of supervised learning algorithm which is used in mail service providers like Gmail, yahoo, etc. to classify a new mail as spam or

not spam.

A : Classification

B : Regression

C : Clustering

D : Naïve bays

Q.no 38. In ---------- the x-axes are grouped into bins and each bin will be treated as a category.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 39. When data are collected in a statistical study for only a portion or subset of all elements of interest we are using

A : Sample

B : Parameter

C : Population

D : Probability

Q.no 40. ---------- regression finds a relaitionship between one or more features (independent variables) and a continuous variables (dependent variable).

A : Non-linear

B : Linear

C : Both of these

D : None of These

Q.no 41. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 42. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Q.no 43. --------- is technique that duplicates smaller array to make dimensionality and size of an array as the size and dimensionality of larger array.

A : Multiplation

B : Broadcasting

C : Addition

D : Flatten

Q.no 44. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 45. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 46. ---------- machine learning algorithm used in cross marketing to work with other businesss that complement your own business but not to other competitors.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 47. Which function returns an ndarray object that contains the numbers that are evenly spaced on a log scale.

A : numpy.logspace()

B : numpy.log()

C : numpy.fill()

D : numpy.random()

Q.no 48. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 49. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 50. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Q.no 51. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 52. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 53. Apriori algorithm uses breadth first search and ----------structure to count candidate item sets efficiently.

A : Decision tree

B : Hash tree

C : Red-Black Tree

D : AVL Tree

Q.no 54. In Data science project data acquisition step involves----------

A : Acquiring data from various sources.

B : Selecting dataset

C : Data preprocessing

D : Data modeling

Q.no 55. Select the correct statement:

A : Raw data is original source of data.

B : Preprocessed data is original source of data.

C : Raw data is the data obtained after processing steps.

D : Analysed data is original source of data.

Q.no 56. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 57. Catelog design is complex process where the selection of items in a business's catelog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related. Which algorith

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 58. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 59. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 60. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Answer for Question No 1. is a

Answer for Question No 2. is c

Answer for Question No 3. is c

Answer for Question No 4. is d

Answer for Question No 5. is c

Answer for Question No 6. is c

Answer for Question No 7. is a

Answer for Question No 8. is c

Answer for Question No 9. is b

Answer for Question No 10. is c

Answer for Question No 11. is a

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is b

Answer for Question No 15. is d

Answer for Question No 16. is b

Answer for Question No 17. is b

Answer for Question No 18. is c

Answer for Question No 19. is a

Answer for Question No 20. is b

Answer for Question No 21. is b

Answer for Question No 22. is a

Answer for Question No 23. is d

Answer for Question No 24. is d

Answer for Question No 25. is c

Answer for Question No 26. is b

Answer for Question No 27. is a

Answer for Question No 28. is c

Answer for Question No 29. is a

Answer for Question No 30. is c

Answer for Question No 31. is b

Answer for Question No 32. is c

Answer for Question No 33. is a

Answer for Question No 34. is a

Answer for Question No 35. is c

Answer for Question No 36. is a

Answer for Question No 37. is a

Answer for Question No 38. is d

Answer for Question No 39. is a

Answer for Question No 40. is b

Answer for Question No 41. is a

Answer for Question No 42. is d

Answer for Question No 43. is b

Answer for Question No 44. is d

Answer for Question No 45. is b

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is d

Answer for Question No 49. is d

Answer for Question No 50. is c

Answer for Question No 51. is a

Answer for Question No 52. is a

Answer for Question No 53. is b

Answer for Question No 54. is a

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is b

Answer for Question No 58. is a

Answer for Question No 59. is b

Answer for Question No 60. is a

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. In statistics, a population consists of ----------

A : All People living in a country.

B : All People living in the city.

C : All subjects or objects whose characteristics are being studied.

D : Part of whole dataset

Q.no 2. ---------- data that depends on data model and resides in a fixed field within a record.

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 3. ---------- plot displays information as series of data points connected by straight lines.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 4. ---------- is about developing code to enable the machine to learn to perform tasks and its basic principle is the automatic modeling of underlying that have generated the collected data.

A : Data Science

B : Data Analytics

C : Data Warehousing

D : Data mining

Q.no 5. The -------- function creates a 2-D array with all values 1.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 6. ---------- method is dataframe reads first n rows from dataframe

A : head(n)

B : tail(n)

C : first(n)

D : start(n)

Q.no 7. Numpy support this function to find trigonometric sine elementwise .

A : numpy.sin()

B : numpy.cosine()

C : numpy.tangent()

D : numpy.rad2sin(x1)

Q.no 8. Apriori algorithm is ---------- machine learning algorithm.

A : Un- Supervised

B : Supervised

C : Both of these

D : None of These

Q.no 9. Which library from python is used for implementing machine learning algorithms?

A : Scikit-Learn

B : Pandas

C : Matplotlib

D : Numpy

Q.no 10. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 11. Which of the following is not a raster image file format?

A : PNG

B : JPG

C : BMP

D : PDF

Q.no 12. K- nearest neighbors algorithm is based on ---------- learning

A : Un- Supervised

B : Supervised

C : Association

D : correlation

Q.no 13. ---------- is an example of human generated unstructured data.

A : YouTube data

B : Satellite data

C : Sensor data

D : Seismic imagery data

Q.no 14. Which of the following is NOT supervised learning?

A : PCA

B : Decision Tree

C : Linear Regression

D : Naive Bayesian

Q.no 15. ---------- is supervised machine learning algorithm outputs an optimal hyperplane for given labled training data

A : KNN

B : Support Vector Machines

C : Regression

D : Decision Tree

Q.no 16. ---------- rule mining is a technique to identify underlying relations between different items.

A : Classification

B : Regression

C : Clustering

D : Association

Q.no 17. ----------type of analytics descibes what happened in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 18. -------- function is used to add a title to each axis instance in a figure.

A : set_title()

B : get_title()

C : set_label()

D : title()

Q.no 19. Which function is used to give title for the axes.

A : plt.title()

B : plt.xlabel()

C : plt.ylabel()

D : plt.xscale()

Q.no 20. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 21. In ---------- the x-axes are grouped into bins and each bin will be treated as a category.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 22. ------- is basic data structure of pandas can be think of SQL table or a spreadsheet data representation.

A : Dataframe

B : series

C : list

D : ndarray

Q.no 23. From matplotlib---------- module is used for plotting various plots.

A : Scilearn

B : Pyplot

C : Scilab

D : Matlab

Q.no 24. A perfect negative correlation is signified by ----------

A : 1

B : -1

C : 0

D : 2

Q.no 25. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 26. In matplotlib library ---------- module supports basic image loading, rescaling and display operations.

A : picture

B : image

C : pyplot

D : sympy

Q.no 27. --------- function from matplotlib.pyplot library plots bar graph for given values of x and y.

A : plot()

B : draw()

C : bar()

D : linedraw()

Q.no 28. ---------- is unsupervised technique aiming to divide a multivariate dataset into clusters or groups.

A : KNN

B : Support Vector Machines

C : Regression

D : Cluster analysis

Q.no 29. When data are collected in a statistical study for only a portion or subset of all elements of interest we are using

A : Sample

B : Parameter

C : Population

D : Probability

Q.no 30. -------- is most important language for Data Science.

A : Java

B : Ruby

C : R

D : None of these

Q.no 31. The last element of ndarray is indexed by ----------

A : 0

B : -1

C : 1

D : -2

Q.no 32. The number of iterations in apriori ----------

A : increases with the size of the data

B : decreases with the increase in size of the data

C : increases with the size of the maximum frequent set

D : decreases with increase in size of the maximum frequent set

Q.no 33. Which of the following is used as attribute selection measure in decision tree algorithms?

A : Information Gain

B : Posterior probability

C : Prior probability

D : Support

Q.no 34. ----------is not one of the key data science skill.

A : Statistics

B : Machine Learning

C : Data Visualization

D : software tester

Q.no 35. What is correct syntax to generate inetegers between 10 to 30

A : x=numpy.arange(10,30)

B : x=numpy.array(10,30)

C : x=numpy.arange(10,31)

D : x=arange(10,31)

Q.no 36. ---------- is unsupervised machine learning technique.

A : KNN

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 37. ---------- searches for the linear optimal separating hyperplane for separation of the data using essential training tuples called support vectors

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 38. ---------- is a one dimensiional array defined in pandas that can be used to store any data type.

A : Dict

B : series

C : ndarray

D : list

Q.no 39. To read image from a file into an array ---------- function is used.

A : matplotlib.pyplot.imshow()

B : matplotlib.pyplot.imread()

C : matplotlib.pyplot.imwrite()

D : matplotlib.pyplot.imsave()

Q.no 40. JSON file data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 41. In regression the independent variable is also called as ----------

A : Regressor

B : Continuous

C : Regressand

D : Estimated

Q.no 42. ---------- function from scipy is used to calculate the distance between all pairs of points in a given set.

A : scipy.spatial.distance()

B : scipy.spatial.distance.measure()

C : scipy.spatial.distance.cdist()

D : distance(x1,y1)

Q.no 43. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 44. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 45. To determine basic salary of a employee when his qualification is given is a ---------- problem

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 46. Which function from numpy used to return the truncated value of the input elementwise?

A : round()

B : trunc()

C : del()

D : remove_decimal()

Q.no 47. Apriori algorithm uses breadth first search and ----------structure to count candidate item sets efficiently.

A : Decision tree

B : Hash tree

C : Red-Black Tree

D : AVL Tree

Q.no 48. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 49. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 50. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 51. The strength (degree) of the correlation between a set of independent variables X and a dependent variable Y is measured by----------

A : Coefficient of Correlation

B : Coefficient of Determination

C : Standard error of estimate

D : Probability

Q.no 52. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 53. When there is no impact on one variable when increse or decrese on other variable then it is ----------

A : Perfect correlation

B : No Correlation

C : Positive Correlation

D : Negative Correlation

Q.no 54. In matplotlib ---------- is container class for figure instance.

A : Axes

B : Canvas

C : Figure

D : FigureCanvas

Q.no 55. Plot_number parameter from subplot() function can range from 1 to ------

A : nrows*ncols

B : max

C : nrows

D : ncols

Q.no 56. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 57. ---------- machine learning algorithm used in cross marketing to work with other businesss that complement your own business but not to other competitors.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 58. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Q.no 59. In dataframe to compute summary statistics like mean, standard deviation, min and max count etc for each numerical column ---------- function is used.

A : display()

B : head()

C : describe()

D : sort()

Q.no 60. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Answer for Question No 1. is c

Answer for Question No 2. is a

Answer for Question No 3. is b

Answer for Question No 4. is b

Answer for Question No 5. is a

Answer for Question No 6. is a

Answer for Question No 7. is a

Answer for Question No 8. is a

Answer for Question No 9. is a

Answer for Question No 10. is d

Answer for Question No 11. is d

Answer for Question No 12. is b

Answer for Question No 13. is a

Answer for Question No 14. is a

Answer for Question No 15. is b

Answer for Question No 16. is d

Answer for Question No 17. is a

Answer for Question No 18. is a

Answer for Question No 19. is a

Answer for Question No 20. is a

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is b

Answer for Question No 24. is c

Answer for Question No 25. is a

Answer for Question No 26. is b

Answer for Question No 27. is c

Answer for Question No 28. is d

Answer for Question No 29. is a

Answer for Question No 30. is c

Answer for Question No 31. is b

Answer for Question No 32. is c

Answer for Question No 33. is a

Answer for Question No 34. is d

Answer for Question No 35. is c

Answer for Question No 36. is d

Answer for Question No 37. is d

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is c

Answer for Question No 41. is a

Answer for Question No 42. is c

Answer for Question No 43. is a

Answer for Question No 44. is d

Answer for Question No 45. is b

Answer for Question No 46. is b

Answer for Question No 47. is b

Answer for Question No 48. is a

Answer for Question No 49. is a

Answer for Question No 50. is a

Answer for Question No 51. is a

Answer for Question No 52. is b

Answer for Question No 53. is b

Answer for Question No 54. is d

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is b

Answer for Question No 58. is b

Answer for Question No 59. is c

Answer for Question No 60. is a

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 2. ---------- means part of population chosen for participation in the study

A : Population

B : Sample

C : Association

D : Correlation

Q.no 3. Choose correct option for machine generated unstructured data.

A : Website data

B : YouTube data

C : Text File data

D : Sensor data

Q.no 4. To save or write dataframe data into csv file -------- function is used

A : write_csv()

B : write_file()

C : csv_read()

D : to_csv()

Q.no 5. ---------- uses a tree structure to specify sequences ofdecisions and consequences.

A : Regression

B : Decision trees

C : KNN

D : SVM

Q.no 6. ---------- is about developing code to enable the machine to learn to perform tasks and its basic principle is the automatic modeling of underlying that have generated the collected data.

A : Data Science

B : Data Analytics

C : Data Warehousing

D : Data mining

Q.no 7. Numpy support this function to find trigonometric sine elementwise .

A : numpy.sin()

B : numpy.cosine()

C : numpy.tangent()

D : numpy.rad2sin(x1)

Q.no 8. ----------type of analytics descibes what happened in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 9. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 10. Sattelite image is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 11. Unsupervised learning makes sense of ---------- data without having any predefined dataset for its training.

A : unlabled

B : labeled

C : semi-labled

D : Empty dataset

Q.no 12. Correlation coefficient values lies between----- and ---

A : -1 and +1

B : -1 and 0

C : 0 and 1

D : 0 and infinite

Q.no 13. K- nearest neighbors algorithm is based on ---------- learning

A : Un- Supervised

B : Supervised

C : Association

D : correlation

Q.no 14. ------ answers the questions like " How can we make it happen?"

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 15. ---------- type of plots show all individual data points without connected with lines.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 16. ---------- chart is a circular plot divides into sclices to show numerical proportion.

A : Bar

B : Line

C : Scatter

D : Pie

Q.no 17. Which of the following is measure used in decision trees while selecting splliting criteria that partitions data into the best possible manner.

A : Information Gain

B : Probability

C : Regression

D : Association

Q.no 18. ---------- is an example of human generated unstructured data.

A : YouTube data

B : Satellite data

C : Sensor data

D : Seismic imagery data

Q.no 19. ---------- charts represents categorical data with retangular bars

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 20. In correlation both values are always----------

A : Random

B : sequential

C : Same

D : from same group

Q.no 21. To rotate an image -------- function is used from scipy library.

A : rotation()

B : scipy.move()

C : scipy.ndimage.rotate()

D : scipy.flip()

Q.no 22. A ---------- is an example of the most widely used machine learning algorithms much of its popularity is because it can be adapted to almost any type od data.

A : Clustering

B : Regression

C : Decision trees

D : Apriori

Q.no 23. ------ is a classification technique relies on the naïve assumption that input variables are independent of each other.

A : KNN

B : NAïve Bayes

C : Regression

D : Support vector machine

Q.no 24. ---------- phase of the data analytics lifecycle usually takes the longest time.

A : Data Preparation

B : Model Planning

C : Model Building

D : Communicate Results

Q.no 25. ---------- is an excellent 2D and 3D graphics library for generating scientific figures?

A : Pandas

B : Numpy

C : matplotlib

D : ndarray

Q.no 26. -------- is most important language for Data Science.

A : Java

B : Ruby

C : R

D : None of these

Q.no 27. Which statement will create 5 x 5 array filled with all values 1

A : x=numpy.ones((5,5))

B : x=numpy.ones(5)

C : x=numpy.zeros((5,5))

D : x=numpy.eye((5,5))

Q.no 28. Which function returns the identity array with n x n dimension with its main diagonal set to ones and all other elements to zero.

A : numpy.ones()

B : numpy.zeros()

C : numpy.fill()

D : numpy.identity()

Q.no 29. From matplotlib---------- module is used for plotting various plots.

A : Scilearn

B : Pyplot

C : Scilab

D : Matlab

Q.no 30. In this type of clustring each data type either belongs to acluster completely or not.

A : Hard clustering

B : Soft Clustering

C : Medium clustering

D : Simple clustring

Q.no 31. ---------- function used to add two numppy arrays elementwise.

A : numpy.add(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.addition(x1,x2)

Q.no 32. A ----------graph is a circular plot, divided into slices to show numerical proportions.

A : Bar

B : Scatter

C : pie

D : line

Q.no 33. --------- function from matplotlib.pyplot library plots bar graph for given values of x and y.

A : plot()

B : draw()

C : bar()

D : linedraw()

Q.no 34. If a=np.array([1,2,3,4,5,6,7,8,9,10]) then a[2,5,1] will produce output----------

A : 3, 4, 5

B : 3,4,5,6

C : 2,3,4,5

D : 1,2,3,4,5

Q.no 35. Identify the machine generated unstructured data.

A : Website data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 36. ----------is not one of the key data science skill.

A : Statistics

B : Machine Learning

C : Data Visualization

D : software tester

Q.no 37. ---------- is raster graphic format with lossless compression.

A : EPS

B : PDF

C : PNG

D : PS

Q.no 38. ------ module from sklearn gathers popular unsupervised clustering algorithms.

A : sklearn.covariance

B : sklearn.base

C : sklearn.neighbors

D : sklearn.cluster

Q.no 39. Regression analysis ----------

A : Establishes a relationship between two variables

B : Establishes cause and effect

C : Measures growth

D : Measures demand for good

Q.no 40. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 41. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 42. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 43. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 44. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 45. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 46. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 47. --------- function performs the custom operations for the entire dataframe.

A : function()

B : surutine()

C : rutine()

D : pipe()

Q.no 48. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 49. Which of the following algorithm is used in Economics, Finance, Biology etc, to model relationships between parameters of intrests.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 50. The statement subplot(4,3,5) will divide figure into ------- and specify plotting sholud be done on plot number----------

A : 4 x 3, 5

B : 3x 4, 5

C : 3 x 5, 4

D : 5x 3, 4

Q.no 51. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 52. ---------- is basically extracting particular set of elements from an array.

A : Slicing

B : indexing

C : sorting

D : broadcasting

Q.no 53. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 54. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Q.no 55. Plot_number parameter from subplot() function can range from 1 to ------

A : nrows*ncols

B : max

C : nrows

D : ncols

Q.no 56. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 57. Catelog design is complex process where the selection of items in a business's catelog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related. Which algorith

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 58. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Q.no 59. In dataframe to compute summary statistics like mean, standard deviation, min and max count etc for each numerical column ---------- function is used.

A : display()

B : head()

C : describe()

D : sort()

Q.no 60. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Answer for Question No 1. is a

Answer for Question No 2. is b

Answer for Question No 3. is d

Answer for Question No 4. is d

Answer for Question No 5. is b

Answer for Question No 6. is b

Answer for Question No 7. is a

Answer for Question No 8. is a

Answer for Question No 9. is d

Answer for Question No 10. is b

Answer for Question No 11. is a

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is b

Answer for Question No 15. is c

Answer for Question No 16. is d

Answer for Question No 17. is a

Answer for Question No 18. is a

Answer for Question No 19. is a

Answer for Question No 20. is a

Answer for Question No 21. is c

Answer for Question No 22. is c

Answer for Question No 23. is b

Answer for Question No 24. is a

Answer for Question No 25. is c

Answer for Question No 26. is c

Answer for Question No 27. is a

Answer for Question No 28. is d

Answer for Question No 29. is b

Answer for Question No 30. is a

Answer for Question No 31. is a

Answer for Question No 32. is c

Answer for Question No 33. is c

Answer for Question No 34. is a

Answer for Question No 35. is d

Answer for Question No 36. is d

Answer for Question No 37. is c

Answer for Question No 38. is d

Answer for Question No 39. is a

Answer for Question No 40. is a

Answer for Question No 41. is d

Answer for Question No 42. is d

Answer for Question No 43. is a

Answer for Question No 44. is a

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is d

Answer for Question No 48. is a

Answer for Question No 49. is a

Answer for Question No 50. is a

Answer for Question No 51. is a

Answer for Question No 52. is a

Answer for Question No 53. is c

Answer for Question No 54. is c

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is b

Answer for Question No 58. is b

Answer for Question No 59. is c

Answer for Question No 60. is d

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Apriori algorithm is ---------- machine learning algorithm.

A : Un- Supervised

B : Supervised

C : Both of these

D : None of These

Q.no 2. CCTV footaage is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 3. Choose correct option for machine generated unstructured data.

A : Website data

B : YouTube data

C : Text File data

D : Sensor data

Q.no 4. Pin code of a city is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 5. The leaf nodes in decision trees returns the ---------

A : decision condition

B : class lables

C : decision on variables

D : test score

Q.no 6. ---------- provides arange of supervised and un-supervised learning algorithms via consistant interface in python

A : Pandas

B : Numpy

C : Scikit-Learn

D : image

Q.no 7. To import data from excel file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : read()

D : read_excel()

Q.no 8. ---------- function used to get positive square root of an numppy array elementwise.

A : numpy.sqrt(x1)

B : numpy.mod(x1)

C : numpy.square(x1)

D : numpy.find(x1,2)

Q.no 9. ----------function reads an image from a file as an array.

A : imsave()

B : imread()

C : read()

D : None of these

Q.no 10. Numpy support this function to find trigonometric sine elementwise .

A : numpy.sin()

B : numpy.cosine()

C : numpy.tangent()

D : numpy.rad2sin(x1)

Q.no 11. In statistics, a population consists of ----------

A : All People living in a country.

B : All People living in the city.

C : All subjects or objects whose characteristics are being studied.

D : Part of whole dataset

Q.no 12. In numpy array , array indices always starts from --------

A : 1

B : -1

C : 0

D : 2

Q.no 13. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 14. ---------- referes to the graphical represetation of information and data.

A : Data Visualization

B : Data mining

C : Data warehousing

D : Data Structures

Q.no 15. ---------- rule mining is a technique to identify underlying relations between different items.

A : Classification

B : Regression

C : Clustering

D : Association

Q.no 16. ---------- means part of population chosen for participation in the study

A : Population

B : Sample

C : Association

D : Correlation

Q.no 17. Email data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 18. Probability always lies between ----- and ----

A : 0 and 1

B : -1 and +1

C : -1 and 0

D : 0 and infinite

Q.no 19. Which of the following is not a type of clustering algorithm?

A : Density clustering

B : K-Mean clustering

C : Centroid clustering

D : Simple clustering

Q.no 20. ---------- plot displays information as series of data points connected by straight lines.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 21. ------ module from sklearn gathers popular unsupervised clustering algorithms.

A : sklearn.covariance

B : sklearn.base

C : sklearn.neighbors

D : sklearn.cluster

Q.no 22. ---------- is an example of semi structured data

A : NoSQL data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 23. Which of the following is used as attribute selection measure in decision tree algorithms?

A : Information Gain

B : Posterior probability

C : Prior probability

D : Support

Q.no 24. A ----------graph is a circular plot, divided into slices to show numerical proportions.

A : Bar

B : Scatter

C : pie

D : line

Q.no 25. ---------- searches for the linear optimal separating hyperplane for separation of the data using essential training tuples called support vectors

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 26. ----------the step is performed by data scientist after acquiring the data.

A : Data Cleansing

B : Data Integration

C : Data Replication

D : Data loading

Q.no 27. Which function returns the identity array with n x n dimension with its main diagonal set to ones and all other elements to zero.

A : numpy.ones()

B : numpy.zeros()

C : numpy.fill()

D : numpy.identity()

Q.no 28. --------- function from matplotlib.pyplot library plots bar graph for given values of x and y.

A : plot()

B : draw()

C : bar()

D : linedraw()

Q.no 29. ---------- is an excellent 2D and 3D graphics library for generating scientific figures?

A : Pandas

B : Numpy

C : matplotlib

D : ndarray

Q.no 30. The process by which we estimate value of dependent variable on the basis of one or more independent variables is called as ----------

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 31. A ---------- is an example of the most widely used machine learning algorithms much of its popularity is because it can be adapted to almost any type od data.

A : Clustering

B : Regression

C : Decision trees

D : Apriori

Q.no 32. Slop of the regression line of Y on X is also called as

A : Correlation coefficient

B : Regression coefficient

C : Association coefficient

D : Probability

Q.no 33. -------- is the measure of the likeihood that an event will occure in a random experiment

A : Probability

B : Correlation

C : Regression

D : Sample

Q.no 34. What is the use of following function? Plt.xlabel("Total Marks")

A : Gives label to X-Axis

B : Gives label to Y-Axis

C : Gives title to figure

D : Add text to figure

Q.no 35. ---------- analysis finds the reasons behind success or failure in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 36. Pandas provide ---------- function as the entry point for all standard database join operations while merging two DataFrame objects.

A : concat()

B : replace()

C : merge()

D : add()

Q.no 37. JSON file data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 38. Broadcasting is a powerful technique that allows numpy to work with arrays of ---------- .

A : Same Shapes

B : Different Shapes

C : Same values

D : Different values

Q.no 39. If scatter diagram is drawn and all scatter points lie on a straight line then it indicates-------

A : No correlation

B : Perfect correlation

C : Regression

D : Skewness

Q.no 40. ---------- models search the data space for areas of varied density of data points in the data space.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 41. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 42. In matplotlib ---------- is container class for figure instance.

A : Axes

B : Canvas

C : Figure

D : FigureCanvas

Q.no 43. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 44. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 45. Catelog design is complex process where the selection of items in a business's catelog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related. Which algorith

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 46. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 47. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Q.no 48. --------- function performs the custom operations for the entire dataframe.

A : function()

B : surutine()

C : rutine()

D : pipe()

Q.no 49. For testing accuracy of a machine learning algorithm whole data set should be devided into trainin and testing datasets. Which of the following is good preportion for train-test spliting?

A : Train- 70%, Test - 30%

B : Train- 50%, Test - 50%

C : Train- 30%, Test - 70%

D : Train- 100%, Test - 00%

Q.no 50. Which function from numpy used to return the truncated value of the input elementwise?

A : round()

B : trunc()

C : del()

D : remove_decimal()

Q.no 51. When there is no impact on one variable when increse or decrese on other variable then it is ----------

A : Perfect correlation

B : No Correlation

C : Positive Correlation

D : Negative Correlation

Q.no 52. Select the correct statement:

A : Raw data is original source of data.

B : Preprocessed data is original source of data.

C : Raw data is the data obtained after processing steps.

D : Analysed data is original source of data.

Q.no 53. --------- is technique that duplicates smaller array to make dimensionality and size of an array as the size and dimensionality of larger array.

A : Multiplation

B : Broadcasting

C : Addition

D : Flatten

Q.no 54. Apriori algorithm uses breadth first search and ----------structure to count candidate item sets efficiently.

A : Decision tree

B : Hash tree

C : Red-Black Tree

D : AVL Tree

Q.no 55. The statement subplot(4,3,5) will divide figure into ------- and specify plotting sholud be done on plot number----------

A : 4 x 3, 5

B : 3x 4, 5

C : 3 x 5, 4

D : 5x 3, 4

Q.no 56. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 57. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Q.no 58. Which function returns an ndarray object that contains the numbers that are evenly spaced on a log scale.

A : numpy.logspace()

B : numpy.log()

C : numpy.fill()

D : numpy.random()

Q.no 59. ---------- function from scipy is used to calculate the distance between all pairs of points in a given set.

A : scipy.spatial.distance()

B : scipy.spatial.distance.measure()

C : scipy.spatial.distance.cdist()

D : distance(x1,y1)

Q.no 60. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Answer for Question No 1. is a

Answer for Question No 2. is b

Answer for Question No 3. is d

Answer for Question No 4. is a

Answer for Question No 5. is b

Answer for Question No 6. is c

Answer for Question No 7. is d

Answer for Question No 8. is a

Answer for Question No 9. is b

Answer for Question No 10. is a

Answer for Question No 11. is c

Answer for Question No 12. is c

Answer for Question No 13. is a

Answer for Question No 14. is a

Answer for Question No 15. is d

Answer for Question No 16. is b

Answer for Question No 17. is b

Answer for Question No 18. is a

Answer for Question No 19. is d

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is a

Answer for Question No 24. is c

Answer for Question No 25. is d

Answer for Question No 26. is a

Answer for Question No 27. is d

Answer for Question No 28. is c

Answer for Question No 29. is c

Answer for Question No 30. is b

Answer for Question No 31. is c

Answer for Question No 32. is b

Answer for Question No 33. is a

Answer for Question No 34. is a

Answer for Question No 35. is a

Answer for Question No 36. is c

Answer for Question No 37. is c

Answer for Question No 38. is b

Answer for Question No 39. is b

Answer for Question No 40. is d

Answer for Question No 41. is b

Answer for Question No 42. is d

Answer for Question No 43. is a

Answer for Question No 44. is a

Answer for Question No 45. is b

Answer for Question No 46. is a

Answer for Question No 47. is c

Answer for Question No 48. is d

Answer for Question No 49. is a

Answer for Question No 50. is b

Answer for Question No 51. is b

Answer for Question No 52. is a

Answer for Question No 53. is b

Answer for Question No 54. is b

Answer for Question No 55. is a

Answer for Question No 56. is d

Answer for Question No 57. is d

Answer for Question No 58. is a

Answer for Question No 59. is c

Answer for Question No 60. is b

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Email data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 2. The procedure to organize items of a given collection into groups based on some similar features called as ----------

A : Regression

B : Clustering

C : Ddecion Trees

D : Association

Q.no 3. ---------- is fundamental library used for scientific computing

A : Pandas

B : Numpy

C : Sympy

D : Scipy

Q.no 4. -------- function is used to add a title to each axis instance in a figure.

A : set_title()

B : get_title()

C : set_label()

D : title()

Q.no 5. ---------- provides arange of supervised and un-supervised learning algorithms via consistant interface in python

A : Pandas

B : Numpy

C : Scikit-Learn

D : image

Q.no 6. The -------- function creates a 2-D array with diagonal values 1 and rest values zeros.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 7. ---------- referes to the graphical represetation of information and data.

A : Data Visualization

B : Data mining

C : Data warehousing

D : Data Structures

Q.no 8. To import data from csv file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : csv_read()

D : Frrom_csv()

Q.no 9. The -------- function creates a 2-D array with all values 1.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 10. Naïve Bayes is a classification technique based on ----------

A : Bayes Theorem

B : Pythagorous Theorom

C : Least square method

D : mean square method

Q.no 11. ---------- means part of population chosen for participation in the study

A : Population

B : Sample

C : Association

D : Correlation

Q.no 12. If number of input features are 3 then optimal hyperplane in support vector machine is ----------

A : Single point

B : Line

C : 2-D Plane

D : Non linear line

Q.no 13. ---------- method is dataframe reads first n rows from dataframe

A : head(n)

B : tail(n)

C : first(n)

D : start(n)

Q.no 14. ---------- uses a tree structure to specify sequences ofdecisions and consequences.

A : Regression

B : Decision trees

C : KNN

D : SVM

Q.no 15. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 16. -------- library is built on the top of Numpy, SciPy and Matplotlib

A : Sympy

B : Scikit

C : Pandas

D : Numpy

Q.no 17. Which library from python is used for implementing machine learning algorithms?

A : Scikit-Learn

B : Pandas

C : Matplotlib

D : Numpy

Q.no 18. ---------- chart is a circular plot divides into sclices to show numerical proportion.

A : Bar

B : Line

C : Scatter

D : Pie

Q.no 19. Sattelite image is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 20. Which of the following is not a raster image file format?

A : PNG

B : JPG

C : BMP

D : PDF

Q.no 21. Which of the following plots is not used for multidimensional visualization?

A : Andrrews Curves

B : Prallel Chart

C : Deviation Chart

D : Bar

Q.no 22. -------- is the measure of the likeihood that an event will occure in a random experiment

A : Probability

B : Correlation

C : Regression

D : Sample

Q.no 23. The ----- algorithm is the simplest machine learning algorithm, which building the model consists only of storing the training dataset. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset i.e its

A : Apriori

B : K-Nearest Neighbors

C : K-Means

D : Decision Trees

Q.no 24. If X and Y are both independent of each other, then correlation coefficient is ---------

A : 1

B : -1

C : 0

D : 2

Q.no 25. To rotate an image -------- function is used from scipy library.

A : rotation()

B : scipy.move()

C : scipy.ndimage.rotate()

D : scipy.flip()

Q.no 26. To set x Axis lable of a figure---------- function is used

A : set_title()

B : set_lable()

C : set_xlabel()

D : get_xlabel()

Q.no 27. In head()/tail()functions of dataframe the default number of elements to display is --------

A : 3

B : 5

C : 1

D : 10

Q.no 28. Regression analysis ----------

A : Establishes a relationship between two variables

B : Establishes cause and effect

C : Measures growth

D : Measures demand for good

Q.no 29. ---------- is an indication of how frequently the itemset appears in the dataset in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 30. In decision trees leaf node denotes a ----------

A : class distribution

B : test on an attribute

C : outcome of the test

D : class labels

Q.no 31. ---------- analysis finds the reasons behind success or failure in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 32. In this type of algorithms inputs are provided but not the desired output.

A : Cluster analysis

B : Support Vector Machines

C : Decision trees

D : Naïve bays

Q.no 33. Pandas provide ---------- function as the entry point for all standard database join operations while merging two DataFrame objects.

A : concat()

B : replace()

C : merge()

D : add()

Q.no 34. ---------- is 2-D data structure defined in pandas in which data arranged in rows and columns.

A : Series

B : Dataframe

C : ndarray

D : list

Q.no 35. ---------- is an example of semi structured data

A : NoSQL data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 36. ----------the step is performed by data scientist after acquiring the data.

A : Data Cleansing

B : Data Integration

C : Data Replication

D : Data loading

Q.no 37. Entropy is a measure of the randomness in the information being processed.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 38. The process by which we estimate value of dependent variable on the basis of one or more independent variables is called as ----------

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 39. ------- is basic data structure of pandas can be think of SQL table or a spreadsheet data representation.

A : Dataframe

B : series

C : list

D : ndarray

Q.no 40. ---------- regression finds a relaitionship between one or more features (independent variables) and a continuous variables (dependent variable).

A : Non-linear

B : Linear

C : Both of these

D : None of These

Q.no 41. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 42. ---------- machine learning algorithm used in cross marketing to work with other businesss that complement your own business but not to other competitors.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 43. In dataframe to compute summary statistics like mean, standard deviation, min and max count etc for each numerical column ---------- function is used.

A : display()

B : head()

C : describe()

D : sort()

Q.no 44. Catelog design is complex process where the selection of items in a business's catelog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related. Which algorith

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 45. For testing accuracy of a machine learning algorithm whole data set should be devided into trainin and testing datasets. Which of the following is good preportion for train-test spliting?

A : Train- 70%, Test - 30%

B : Train- 50%, Test - 50%

C : Train- 30%, Test - 70%

D : Train- 100%, Test - 00%

Q.no 46. ---------- is basically extracting particular set of elements from an array.

A : Slicing

B : indexing

C : sorting

D : broadcasting

Q.no 47. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 48. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 49. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 50. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 51. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 52. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 53. Plot_number parameter from subplot() function can range from 1 to ------

A : nrows*ncols

B : max

C : nrows

D : ncols

Q.no 54. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 55. The statement subplot(4,3,5) will divide figure into ------- and specify plotting sholud be done on plot number----------

A : 4 x 3, 5

B : 3x 4, 5

C : 3 x 5, 4

D : 5x 3, 4

Q.no 56. The strength (degree) of the correlation between a set of independent variables X and a dependent variable Y is measured by----------

A : Coefficient of Correlation

B : Coefficient of Determination

C : Standard error of estimate

D : Probability

Q.no 57. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 58. In matplotlib ---------- is container class for figure instance.

A : Axes

B : Canvas

C : Figure

D : FigureCanvas

Q.no 59. Which of the following machine learning algorithm is used for maret basket analysis means to analyze the association of purchased items in asingle basket or single purchase.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 60. To determine basic salary of a employee when his qualification is given is a ---------- problem

A : Correlation

B : Regression

C : Association

D : Qualitative

Answer for Question No 1. is b

Answer for Question No 2. is b

Answer for Question No 3. is d

Answer for Question No 4. is a

Answer for Question No 5. is c

Answer for Question No 6. is c

Answer for Question No 7. is a

Answer for Question No 8. is a

Answer for Question No 9. is a

Answer for Question No 10. is a

Answer for Question No 11. is b

Answer for Question No 12. is c

Answer for Question No 13. is a

Answer for Question No 14. is b

Answer for Question No 15. is a

Answer for Question No 16. is b

Answer for Question No 17. is a

Answer for Question No 18. is d

Answer for Question No 19. is b

Answer for Question No 20. is d

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is b

Answer for Question No 24. is b

Answer for Question No 25. is c

Answer for Question No 26. is c

Answer for Question No 27. is b

Answer for Question No 28. is a

Answer for Question No 29. is b

Answer for Question No 30. is c

Answer for Question No 31. is a

Answer for Question No 32. is a

Answer for Question No 33. is c

Answer for Question No 34. is b

Answer for Question No 35. is a

Answer for Question No 36. is a

Answer for Question No 37. is a

Answer for Question No 38. is b

Answer for Question No 39. is a

Answer for Question No 40. is b

Answer for Question No 41. is d

Answer for Question No 42. is b

Answer for Question No 43. is c

Answer for Question No 44. is b

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is a

Answer for Question No 48. is b

Answer for Question No 49. is a

Answer for Question No 50. is a

Answer for Question No 51. is d

Answer for Question No 52. is b

Answer for Question No 53. is a

Answer for Question No 54. is a

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is c

Answer for Question No 58. is d

Answer for Question No 59. is b

Answer for Question No 60. is b

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Numpy support this function to find trigonometric sine elementwise .

A : numpy.sin()

B : numpy.cosine()

C : numpy.tangent()

D : numpy.rad2sin(x1)

Q.no 2. SQL record is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 3. ---------- function used to get positive square root of an numppy array elementwise.

A : numpy.sqrt(x1)

B : numpy.mod(x1)

C : numpy.square(x1)

D : numpy.find(x1,2)

Q.no 4. ---------- data does not fits into a data model due to variatins in contents.

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 5. Which of the following is NOT supervised learning?

A : PCA

B : Decision Tree

C : Linear Regression

D : Naive Bayesian

Q.no 6. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 7. Which of the following function is used to create an array of specified shape but filled with random values.

A : numpy.random.ran()

B : rank

C : random.fill()

D : numpy.fillrandom()

Q.no 8. ---------- is an example of human generated unstructured data.

A : YouTube data

B : Satellite data

C : Sensor data

D : Seismic imagery data

Q.no 9. The -------- function creates a 2-D array with all values 1.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 10. The -------- function creates a 2-D array with all values 0 (zeros).

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 11. ---------- is fundamental library used for scientific computing

A : Pandas

B : Numpy

C : Sympy

D : Scipy

Q.no 12. The -------- function creates a 2-D array with diagonal values 1 and rest values zeros.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 13. Pandas provide ---------- method in order to get label based indexing.

A : iloc()

B : loc()

C : ix()

D : xloc()

Q.no 14. The ---------- attribute specifies the number of dimensions or axes of the array.

A : ndarray.size

B : ndarray.dtype

C : ndarray.ndim

D : ndarray.axes

Q.no 15. In support vector machines if input features are 2 then the decision boundries or hyperplane is ----------.

A : 2-D plane

B : 3-D plane

C : Line

D : point

Q.no 16. ----------type of analytics descibes what happened in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 17. ---- is an technique to learn from examples and experience, without being explicitly programmed.

A : Machine Learning

B : Software Testing

C : Computer Science

D : Data mining

Q.no 18. ---------- means part of population chosen for participation in the study

A : Population

B : Sample

C : Association

D : Correlation

Q.no 19. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 20. ---------- chart is a circular plot divides into sclices to show numerical proportion.

A : Bar

B : Line

C : Scatter

D : Pie

Q.no 21. ----------is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 22. What is correct syntax to generate inetegers between 10 to 30

A : x=numpy.arange(10,30)

B : x=numpy.array(10,30)

C : x=numpy.arange(10,31)

D : x=arange(10,31)

Q.no 23. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 24. ---------- function is used to save an array as in image file.

A : matplotlib.pyplot.image()

B : matplotlib.pyplot.imread()

C : matplotlib.pyplot.imwrite()

D : matplotlib.pyplot.imsave()

Q.no 25. If X and Y are both independent of each other, then correlation coefficient is ---------

A : 1

B : -1

C : 0

D : 2

Q.no 26. JSON file data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 27. What is the use of following function? Plt.xlabel("Total Marks")

A : Gives label to X-Axis

B : Gives label to Y-Axis

C : Gives title to figure

D : Add text to figure

Q.no 28. Regression analysis ----------

A : Establishes a relationship between two variables

B : Establishes cause and effect

C : Measures growth

D : Measures demand for good

Q.no 29. In this type of algorithms inputs are provided but not the desired output.

A : Cluster analysis

B : Support Vector Machines

C : Decision trees

D : Naïve bays

Q.no 30. ---------- analysis finds the reasons behind success or failure in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 31. ---------- models search the data space for areas of varied density of data points in the data space.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 32. ---------- function used to get arrays elementwise remainder of division

A : numpy.divide(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.reminder(x1,x2)

Q.no 33. If a=np.array([1,2,3,4,5,6,7,8,9,10]) then a[2,5,1] will produce output----------

A : 3, 4, 5

B : 3,4,5,6

C : 2,3,4,5

D : 1,2,3,4,5

Q.no 34. Slop of the regression line of Y on X is also called as

A : Correlation coefficient

B : Regression coefficient

C : Association coefficient

D : Probability

Q.no 35. The process by which we estimate value of dependent variable on the basis of one or more independent variables is called as ----------

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 36. In head()/tail()functions of dataframe the default number of elements to display is --------

A : 3

B : 5

C : 1

D : 10

Q.no 37. A perfect negative correlation is signified by ----------

A : 1

B : -1

C : 0

D : 2

Q.no 38. ---------- is unsupervised technique aiming to divide a multivariate dataset into clusters or groups.

A : KNN

B : Support Vector Machines

C : Regression

D : Cluster analysis

Q.no 39. Among the following clustering algorithm types in which of the following type the notion of similarity is derived by the closeness of a data point to the centroid of the clusters.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 40. ---------- is an example of semi structured data

A : XML data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 41. Plot_number parameter from subplot() function can range from 1 to ------

A : nrows*ncols

B : max

C : nrows

D : ncols

Q.no 42. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 43. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 44. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 45. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 46. To determine basic salary of a employee when his qualification is given is a ---------- problem

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 47. In Data science project data acquisition step involves----------

A : Acquiring data from various sources.

B : Selecting dataset

C : Data preprocessing

D : Data modeling

Q.no 48. --------- is technique that duplicates smaller array to make dimensionality and size of an array as the size and dimensionality of larger array.

A : Multiplation

B : Broadcasting

C : Addition

D : Flatten

Q.no 49. Which function from numpy used to return the truncated value of the input elementwise?

A : round()

B : trunc()

C : del()

D : remove_decimal()

Q.no 50. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Q.no 51. Which of the following machine learning algorithm is used for maret basket analysis means to analyze the association of purchased items in asingle basket or single purchase.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 52. ---------- machine learning algorithm used in cross marketing to work with other businesss that complement your own business but not to other competitors.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 53. In regression the independent variable is also called as ----------

A : Regressor

B : Continuous

C : Regressand

D : Estimated

Q.no 54. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 55. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 56. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 57. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 58. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 59. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 60. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Answer for Question No 1. is a

Answer for Question No 2. is a

Answer for Question No 3. is a

Answer for Question No 4. is b

Answer for Question No 5. is a

Answer for Question No 6. is a

Answer for Question No 7. is a

Answer for Question No 8. is a

Answer for Question No 9. is a

Answer for Question No 10. is b

Answer for Question No 11. is d

Answer for Question No 12. is c

Answer for Question No 13. is b

Answer for Question No 14. is c

Answer for Question No 15. is c

Answer for Question No 16. is a

Answer for Question No 17. is a

Answer for Question No 18. is b

Answer for Question No 19. is d

Answer for Question No 20. is d

Answer for Question No 21. is a

Answer for Question No 22. is c

Answer for Question No 23. is a

Answer for Question No 24. is d

Answer for Question No 25. is b

Answer for Question No 26. is c

Answer for Question No 27. is a

Answer for Question No 28. is a

Answer for Question No 29. is a

Answer for Question No 30. is a

Answer for Question No 31. is d

Answer for Question No 32. is b

Answer for Question No 33. is a

Answer for Question No 34. is b

Answer for Question No 35. is b

Answer for Question No 36. is b

Answer for Question No 37. is c

Answer for Question No 38. is d

Answer for Question No 39. is b

Answer for Question No 40. is a

Answer for Question No 41. is a

Answer for Question No 42. is a

Answer for Question No 43. is d

Answer for Question No 44. is a

Answer for Question No 45. is a

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is b

Answer for Question No 49. is b

Answer for Question No 50. is c

Answer for Question No 51. is b

Answer for Question No 52. is b

Answer for Question No 53. is a

Answer for Question No 54. is b

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is d

Answer for Question No 58. is c

Answer for Question No 59. is b

Answer for Question No 60. is d

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Unsupervised learning makes sense of ---------- data without having any predefined dataset for its training.

A : unlabled

B : labeled

C : semi-labled

D : Empty dataset

Q.no 2. For multidimensional visualization ---------- are used.

A : pia charts

B : Bar charts

C : Andrews curves

D : Scatter plots

Q.no 3. ---------- referes to the graphical represetation of information and data.

A : Data Visualization

B : Data mining

C : Data warehousing

D : Data Structures

Q.no 4. ---------- function multiply two matrices in numpy.

A : prod()

B : mult()

C : dot()

D : *

Q.no 5. If number of input features are 3 then optimal hyperplane in support vector machine is ----------

A : Single point

B : Line

C : 2-D Plane

D : Non linear line

Q.no 6. Probability always lies between ----- and ----

A : 0 and 1

B : -1 and +1

C : -1 and 0

D : 0 and infinite

Q.no 7. ------ answers the questions like " How can we make it happen?"

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 8. Pandas provide ---------- method in order to get label based indexing.

A : iloc()

B : loc()

C : ix()

D : xloc()

Q.no 9. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 10. ---------- is a general purpose array-processing package provides a high performance multi-dimentional array object and tools for working with these arrays.

A : NumPy

B : SciPy

C : sklearn

D : None of these

Q.no 11. The leaf nodes in decision trees returns the ---------

A : decision condition

B : class lables

C : decision on variables

D : test score

Q.no 12. Sattelite image is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 13. The -------- function creates a 2-D array with all values 0 (zeros).

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 14. ---------- function used to get positive square root of an numppy array elementwise.

A : numpy.sqrt(x1)

B : numpy.mod(x1)

C : numpy.square(x1)

D : numpy.find(x1,2)

Q.no 15. Pin code of a city is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 16. ---------- is fundamental library used for scientific computing

A : Pandas

B : Numpy

C : Sympy

D : Scipy

Q.no 17. Find odd one out from the following :

A : KNN

B : NAïve Bayes

C : Decision Trees

D : Cluster analysis

Q.no 18. ---------- is supervised machine learning algorithm outputs an optimal hyperplane for given labled training data

A : KNN

B : Support Vector Machines

C : Regression

D : Decision Tree

Q.no 19. To import data from csv file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : csv_read()

D : Frrom_csv()

Q.no 20. SQL record is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 21. JSON file data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 22. -------- is most important language for Data Science.

A : Java

B : Ruby

C : R

D : None of these

Q.no 23. ---------- is 2-D data structure defined in pandas in which data arranged in rows and columns.

A : Series

B : Dataframe

C : ndarray

D : list

Q.no 24. ----------is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 25. Which of the following is not used for 2-D Visualisation?

A : pia charts

B : Bar charts

C : Andrews curves

D : Scatter plots

Q.no 26. The -------- of a numpy array is a tuple of integers giving the size of the array along each dimension.

A : axes

B : rank

C : shape

D : size

Q.no 27. Pandas provide ---------- method in order to get purly integer based indexing.

A : iloc()

B : loc()

C : ix()

D : xloc()

Q.no 28. --------- in decision tree measures how much information a feature gives us about the class

A : Information Gain

B : Posterior probability

C : Prior probability

D : probability

Q.no 29. The process by which we estimate value of dependent variable on the basis of one or more independent variables is called as ----------

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 30. ------ module from sklearn gathers popular unsupervised clustering algorithms.

A : sklearn.covariance

B : sklearn.base

C : sklearn.neighbors

D : sklearn.cluster

Q.no 31. A ---------- is a supervised machine learning algorithm which relies on the assumptiion of feature independent to classify input data.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 32. ---------- is a form of supervised learning algorithm which is used in mail service providers like Gmail, yahoo, etc. to classify a new mail as spam or not spam.

A : Classification

B : Regression

C : Clustering

D : Naïve bays

Q.no 33. The objective of --------- algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

A : KNN

B : Support Vector Machines

C : Regression

D : Decision Tree

Q.no 34. --------- function from matplotlib.pyplot library plots bar graph for given values of x and y.

A : plot()

B : draw()

C : bar()

D : linedraw()

Q.no 35. ----------is not one of the key data science skill.

A : Statistics

B : Machine Learning

C : Data Visualization

D : software tester

Q.no 36. In matplotlib ---------- function groups smaller axes that can exist togather within a single figure.

A : subplot()

B : divide_figure()

C : add_fig()

D : group_fig()

Q.no 37. ---------- function is used to save an array as in image file.

A : matplotlib.pyplot.image()

B : matplotlib.pyplot.imread()

C : matplotlib.pyplot.imwrite()

D : matplotlib.pyplot.imsave()

Q.no 38. Entropy is a measure of the randomness in the information being processed.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 39. ---------- function used to add two numppy arrays elementwise.

A : numpy.add(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.addition(x1,x2)

Q.no 40. In this type of clustring each data type either belongs to acluster completely or not.

A : Hard clustering

B : Soft Clustering

C : Medium clustering

D : Simple clustring

Q.no 41. The statement subplot(4,3,5) will divide figure into ------- and specify plotting sholud be done on plot number----------

A : 4 x 3, 5

B : 3x 4, 5

C : 3 x 5, 4

D : 5x 3, 4

Q.no 42. Select the correct statement:

A : Raw data is original source of data.

B : Preprocessed data is original source of data.

C : Raw data is the data obtained after processing steps.

D : Analysed data is original source of data.

Q.no 43. Which function from numpy used to return the truncated value of the input elementwise?

A : round()

B : trunc()

C : del()

D : remove_decimal()

Q.no 44. Which function returns an ndarray object that contains the numbers that are evenly spaced on a log scale.

A : numpy.logspace()

B : numpy.log()

C : numpy.fill()

D : numpy.random()

Q.no 45. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 46. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Q.no 47. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 48. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 49. --------- function performs the custom operations for the entire dataframe.

A : function()

B : surutine()

C : rutine()

D : pipe()

Q.no 50. ---------- is basically extracting particular set of elements from an array.

A : Slicing

B : indexing

C : sorting

D : broadcasting

Q.no 51. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 52. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Q.no 53. Which of the following machine learning algorithm is used for maret basket analysis means to analyze the association of purchased items in asingle basket or single purchase.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 54. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 55. In matplotlib ---------- is container class for figure instance.

A : Axes

B : Canvas

C : Figure

D : FigureCanvas

Q.no 56. Which of the following algorithm is used in Economics, Finance, Biology etc, to model relationships between parameters of intrests.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 57. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 58. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 59. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 60. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Answer for Question No 1. is a

Answer for Question No 2. is c

Answer for Question No 3. is a

Answer for Question No 4. is c

Answer for Question No 5. is c

Answer for Question No 6. is a

Answer for Question No 7. is b

Answer for Question No 8. is b

Answer for Question No 9. is a

Answer for Question No 10. is a

Answer for Question No 11. is b

Answer for Question No 12. is b

Answer for Question No 13. is b

Answer for Question No 14. is a

Answer for Question No 15. is a

Answer for Question No 16. is d

Answer for Question No 17. is d

Answer for Question No 18. is b

Answer for Question No 19. is a

Answer for Question No 20. is a

Answer for Question No 21. is c

Answer for Question No 22. is c

Answer for Question No 23. is b

Answer for Question No 24. is a

Answer for Question No 25. is c

Answer for Question No 26. is c

Answer for Question No 27. is a

Answer for Question No 28. is a

Answer for Question No 29. is b

Answer for Question No 30. is d

Answer for Question No 31. is c

Answer for Question No 32. is a

Answer for Question No 33. is b

Answer for Question No 34. is c

Answer for Question No 35. is d

Answer for Question No 36. is a

Answer for Question No 37. is d

Answer for Question No 38. is a

Answer for Question No 39. is a

Answer for Question No 40. is a

Answer for Question No 41. is a

Answer for Question No 42. is a

Answer for Question No 43. is b

Answer for Question No 44. is a

Answer for Question No 45. is a

Answer for Question No 46. is c

Answer for Question No 47. is b

Answer for Question No 48. is d

Answer for Question No 49. is d

Answer for Question No 50. is a

Answer for Question No 51. is a

Answer for Question No 52. is d

Answer for Question No 53. is b

Answer for Question No 54. is d

Answer for Question No 55. is d

Answer for Question No 56. is a

Answer for Question No 57. is c

Answer for Question No 58. is a

Answer for Question No 59. is b

Answer for Question No 60. is b

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Naïve Bayes is a classification technique based on ----------

A : Bayes Theorem

B : Pythagorous Theorom

C : Least square method

D : mean square method

Q.no 2. ---------- function is used to plot a histogram using matplotlib library

A : hist()

B : bar()

C : pie()

D : scatter()

Q.no 3. ---------- rule mining is a technique to identify underlying relations between different items.

A : Classification

B : Regression

C : Clustering

D : Association

Q.no 4. Probability always lies between ----- and ----

A : 0 and 1

B : -1 and +1

C : -1 and 0

D : 0 and infinite

Q.no 5. To import data from excel file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : read()

D : read_excel()

Q.no 6. In numpy array , array indices always starts from --------

A : 1

B : -1

C : 0

D : 2

Q.no 7. Email data is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 8. ---------- function used to get positive square root of an numppy array elementwise.

A : numpy.sqrt(x1)

B : numpy.mod(x1)

C : numpy.square(x1)

D : numpy.find(x1,2)

Q.no 9. In --------- learning the training is controlled by an external supervisor or teacher.

A : Un- Supervised

B : Supervised

C : semi-supervied

D : group

Q.no 10. For multidimensional visualization ---------- are used.

A : pia charts

B : Bar charts

C : Andrews curves

D : Scatter plots

Q.no 11. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 12. To import data from csv file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : csv_read()

D : Frrom_csv()

Q.no 13. The -------- function creates a 2-D array with all values 1.

A : numpy. Ones()

B : numpy.zeros()

C : numpy.eye()

D : numpy.empty()

Q.no 14. K- nearest neighbors algorithm is based on ---------- learning

A : Un- Supervised

B : Supervised

C : Association

D : correlation

Q.no 15. In support vector machines if input features are 2 then the decision boundries or hyperplane is ----------.

A : 2-D plane

B : 3-D plane

C : Line

D : point

Q.no 16. ---------- submodule of scipy is dedicated to image processing.

A : ndarray

B : spatial

C : ndimage

D : special

Q.no 17. ---------- uses a tree structure to specify sequences ofdecisions and consequences.

A : Regression

B : Decision trees

C : KNN

D : SVM

Q.no 18. Numpy support this function to find trigonometric sine elementwise .

A : numpy.sin()

B : numpy.cosine()

C : numpy.tangent()

D : numpy.rad2sin(x1)

Q.no 19. The procedure to organize items of a given collection into groups based on some similar features called as ----------

A : Regression

B : Clustering

C : Ddecion Trees

D : Association

Q.no 20. matplotlib.pyplot.imread() function is used to ----------

A : save image

B : read image

C : copy image

D : show image

Q.no 21. ---------- models search the data space for areas of varied density of data points in the data space.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 22. Pandas provide ---------- method in order to get purly integer based indexing.

A : iloc()

B : loc()

C : ix()

D : xloc()

Q.no 23. To rotate an image -------- function is used from scipy library.

A : rotation()

B : scipy.move()

C : scipy.ndimage.rotate()

D : scipy.flip()

Q.no 24. ---------- is unsupervised machine learning technique.

A : KNN

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 25. ----------is not one of the key data science skill.

A : Statistics

B : Machine Learning

C : Data Visualization

D : software tester

Q.no 26. The number of iterations in apriori ----------

A : increases with the size of the data

B : decreases with the increase in size of the data

C : increases with the size of the maximum frequent set

D : decreases with increase in size of the maximum frequent set

Q.no 27. ---------- regression finds a relaitionship between one or more features (independent variables) and a continuous variables (dependent variable).

A : Non-linear

B : Linear

C : Both of these

D : None of These

Q.no 28. ----------is a flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 29. Which of the following is not used for 2-D Visualisation?

A : pia charts

B : Bar charts

C : Andrews curves

D : Scatter plots

Q.no 30. Support(B) =

A : (Transacions containing (B)) / (Total Transactions)

B : (Transacions containing (B)) / 100

C : (Total Transactions) / (Transacions containing (B))

D : 100/ (Transacions containing (B))

Q.no 31. In decision trees leaf node denotes a ----------

A : class distribution

B : test on an attribute

C : outcome of the test

D : class labels

Q.no 32. Which of the following is used as attribute selection measure in decision tree algorithms?

A : Information Gain

B : Posterior probability

C : Prior probability

D : Support

Q.no 33. A ---------- is a supervised machine learning algorithm which relies on the assumptiion of feature independent to classify input data.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 34. ---------- function used to get arrays elementwise remainder of division

A : numpy.divide(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.reminder(x1,x2)

Q.no 35. In this type of algorithms inputs are provided but not the desired output.

A : Cluster analysis

B : Support Vector Machines

C : Decision trees

D : Naïve bays

Q.no 36. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 37. --------- function from matplotlib.pyplot library plots bar graph for given values of x and y.

A : plot()

B : draw()

C : bar()

D : linedraw()

Q.no 38. To set x Axis lable of a figure---------- function is used

A : set_title()

B : set_lable()

C : set_xlabel()

D : get_xlabel()

Q.no 39. What is the use of following function? Plt.xlabel("Total Marks")

A : Gives label to X-Axis

B : Gives label to Y-Axis

C : Gives title to figure

D : Add text to figure

Q.no 40. In SciPy ---------- submodule is dedicated to image processing.

A : ndimage

B : ndarray

C : signal

D : io

Q.no 41. Apriori algorithm uses breadth first search and ----------structure to count candidate item sets efficiently.

A : Decision tree

B : Hash tree

C : Red-Black Tree

D : AVL Tree

Q.no 42. Which of the following task is not performed by Data Scientist.

A : Define the question

B : Create reproducible code

C : Challenge results

D : Staff Recruitement

Q.no 43. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 44. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 45. In regression the independent variable is also called as ----------

A : Regressor

B : Continuous

C : Regressand

D : Estimated

Q.no 46. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Q.no 47. Select the correct statement:

A : Raw data is original source of data.

B : Preprocessed data is original source of data.

C : Raw data is the data obtained after processing steps.

D : Analysed data is original source of data.

Q.no 48. When there is no impact on one variable when increse or decrese on other variable then it is ----------

A : Perfect correlation

B : No Correlation

C : Positive Correlation

D : Negative Correlation

Q.no 49. For testing accuracy of a machine learning algorithm whole data set should be devided into trainin and testing datasets. Which of the following is good preportion for train-test spliting?

A : Train- 70%, Test - 30%

B : Train- 50%, Test - 50%

C : Train- 30%, Test - 70%

D : Train- 100%, Test - 00%

Q.no 50. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 51. Plot_number parameter from subplot() function can range from 1 to ------

A : nrows*ncols

B : max

C : nrows

D : ncols

Q.no 52. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 53. ---------- function from scipy is used to calculate the distance between all pairs of points in a given set.

A : scipy.spatial.distance()

B : scipy.spatial.distance.measure()

C : scipy.spatial.distance.cdist()

D : distance(x1,y1)

Q.no 54. In this type of clustring instead of putting each data point into a separate cluster a probability or likelihood of that data point to be in those clusters is assigned.

A : Hard clustering

B : Soft Clustering

C : Medium clustering

D : Simple clustring

Q.no 55. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 56. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 57. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 58. Catelog design is complex process where the selection of items in a business's catelog are often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related. Which algorith

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 59. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 60. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Answer for Question No 1. is a

Answer for Question No 2. is a

Answer for Question No 3. is d

Answer for Question No 4. is a

Answer for Question No 5. is d

Answer for Question No 6. is c

Answer for Question No 7. is b

Answer for Question No 8. is a

Answer for Question No 9. is b

Answer for Question No 10. is c

Answer for Question No 11. is d

Answer for Question No 12. is a

Answer for Question No 13. is a

Answer for Question No 14. is b

Answer for Question No 15. is c

Answer for Question No 16. is c

Answer for Question No 17. is b

Answer for Question No 18. is a

Answer for Question No 19. is b

Answer for Question No 20. is b

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is c

Answer for Question No 24. is d

Answer for Question No 25. is d

Answer for Question No 26. is c

Answer for Question No 27. is b

Answer for Question No 28. is a

Answer for Question No 29. is c

Answer for Question No 30. is a

Answer for Question No 31. is c

Answer for Question No 32. is a

Answer for Question No 33. is c

Answer for Question No 34. is b

Answer for Question No 35. is a

Answer for Question No 36. is a

Answer for Question No 37. is c

Answer for Question No 38. is c

Answer for Question No 39. is a

Answer for Question No 40. is a

Answer for Question No 41. is b

Answer for Question No 42. is d

Answer for Question No 43. is a

Answer for Question No 44. is a

Answer for Question No 45. is a

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is b

Answer for Question No 49. is a

Answer for Question No 50. is a

Answer for Question No 51. is a

Answer for Question No 52. is b

Answer for Question No 53. is c

Answer for Question No 54. is b

Answer for Question No 55. is c

Answer for Question No 56. is d

Answer for Question No 57. is a

Answer for Question No 58. is b

Answer for Question No 59. is d

Answer for Question No 60. is b

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Correlation coefficient values lies between----- and ---

A : -1 and +1

B : -1 and 0

C : 0 and 1

D : 0 and infinite

Q.no 2. ----------type of analytics descibes what happened in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 3. In statistics, a population consists of ----------

A : All People living in a country.

B : All People living in the city.

C : All subjects or objects whose characteristics are being studied.

D : Part of whole dataset

Q.no 4. SQL record is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 5. ----------function reads an image from a file as an array.

A : imsave()

B : imread()

C : read()

D : None of these

Q.no 6. Find odd one out from the following :

A : KNN

B : NAïve Bayes

C : Decision Trees

D : Cluster analysis

Q.no 7. The ---------- algorithm is based on the fact that the algorithm uses prior knowledge to find frequent item set.

A : Clustring

B : Regression

C : Naïve Bays

D : Apriori

Q.no 8. Pin code of a city is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 9. matplotlib.pyplot.imread() function is used to ----------

A : save image

B : read image

C : copy image

D : show image

Q.no 10. Choose correct option for machine generated unstructured data.

A : Website data

B : YouTube data

C : Text File data

D : Sensor data

Q.no 11. Which function is used to give title for the axes.

A : plt.title()

B : plt.xlabel()

C : plt.ylabel()

D : plt.xscale()

Q.no 12. Which of the following is measure used in decision trees while selecting splliting criteria that partitions data into the best possible manner.

A : Information Gain

B : Probability

C : Regression

D : Association

Q.no 13. ---------- means part of population chosen for participation in the study

A : Population

B : Sample

C : Association

D : Correlation

Q.no 14. ---------- is an example of human generated unstructured data.

A : YouTube data

B : Satellite data

C : Sensor data

D : Seismic imagery data

Q.no 15. ---------- function is used to save image into an ndarray.

A : imsave()

B : imread()

C : save()

D : isave()

Q.no 16. ---------- chart is a circular plot divides into sclices to show numerical proportion.

A : Bar

B : Line

C : Scatter

D : Pie

Q.no 17. ------- answers the question "What will happen in future?"

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 18. ---------- method is dataframe reads first n rows from dataframe

A : head(n)

B : tail(n)

C : first(n)

D : start(n)

Q.no 19. ---------- referes to the graphical represetation of information and data.

A : Data Visualization

B : Data mining

C : Data warehousing

D : Data Structures

Q.no 20. ---------- is a general purpose array-processing package provides a high performance multi-dimentional array object and tools for working with these arrays.

A : NumPy

B : SciPy

C : sklearn

D : None of these

Q.no 21. -------- is uses a tree structure to specify sequence of decisions and consequences.

A : KNN

B : NAïve Bayes

C : Regression

D : Decision Tree

Q.no 22. Which statement will create 5 x 5 array filled with all values 1

A : x=numpy.ones((5,5))

B : x=numpy.ones(5)

C : x=numpy.zeros((5,5))

D : x=numpy.eye((5,5))

Q.no 23. In matplotlib library ---------- module supports basic image loading, rescaling and display operations.

A : picture

B : image

C : pyplot

D : sympy

Q.no 24. ---------- function used to get arrays elementwise remainder of division

A : numpy.divide(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.reminder(x1,x2)

Q.no 25. In ---------- the x-axes are grouped into bins and each bin will be treated as a category.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 26. -------- is most important language for Data Science.

A : Java

B : Ruby

C : R

D : None of these

Q.no 27. The ----- algorithm is the simplest machine learning algorithm, which building the model consists only of storing the training dataset. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset i.e its

A : Apriori

B : K-Nearest Neighbors

C : K-Means

D : Decision Trees

Q.no 28. From matplotlib---------- module is used for plotting various plots.

A : Scilearn

B : Pyplot

C : Scilab

D : Matlab

Q.no 29. Among the following clustering algorithm types in which of the following type the notion of similarity is derived by the closeness of a data point to the centroid of the clusters.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 30. ---------- is a form of supervised learning algorithm which is used in mail service providers like Gmail, yahoo, etc. to classify a new mail as spam or not spam.

A : Classification

B : Regression

C : Clustering

D : Naïve bays

Q.no 31. The number of iterations in apriori ----------

A : increases with the size of the data

B : decreases with the increase in size of the data

C : increases with the size of the maximum frequent set

D : decreases with increase in size of the maximum frequent set

Q.no 32. In this type of algorithms inputs are provided but not the desired output.

A : Cluster analysis

B : Support Vector Machines

C : Decision trees

D : Naïve bays

Q.no 33. The objective of --------- algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

A : KNN

B : Support Vector Machines

C : Regression

D : Decision Tree

Q.no 34. Which of the following is used as attribute selection measure in decision tree algorithms?

A : Information Gain

B : Posterior probability

C : Prior probability

D : Support

Q.no 35. ---------- analysis finds the reasons behind success or failure in past

A : Descriptive

B : Prescriptive

C : Predictive

D : Probability

Q.no 36. A ----------graph is a circular plot, divided into slices to show numerical proportions.

A : Bar

B : Scatter

C : pie

D : line

Q.no 37. Support(B) =

A : (Transacions containing (B)) / (Total Transactions)

B : (Transacions containing (B)) / 100

C : (Total Transactions) / (Transacions containing (B))

D : 100/ (Transacions containing (B))

Q.no 38. ----------is not one of the key data science skill.

A : Statistics

B : Machine Learning

C : Data Visualization

D : software tester

Q.no 39. ---------- is an indication of how frequently the itemset appears in the dataset in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 40. When data are collected in a statistical study for only a portion or subset of all elements of interest we are using

A : Sample

B : Parameter

C : Population

D : Probability

Q.no 41. In Data science project data acquisition step involves----------

A : Acquiring data from various sources.

B : Selecting dataset

C : Data preprocessing

D : Data modeling

Q.no 42. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Q.no 43. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 44. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 45. Which function returns an ndarray object that contains the numbers that are evenly spaced on a log scale.

A : numpy.logspace()

B : numpy.log()

C : numpy.fill()

D : numpy.random()

Q.no 46. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 47. -------- is an unsupervised algorithm used for frequent itemset mining.

A : Apriori

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 48. Which function from numpy used to return the truncated value of the input elementwise?

A : round()

B : trunc()

C : del()

D : remove_decimal()

Q.no 49. The strength (degree) of the correlation between a set of independent variables X and a dependent variable Y is measured by----------

A : Coefficient of Correlation

B : Coefficient of Determination

C : Standard error of estimate

D : Probability

Q.no 50. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Q.no 51. Which of the following statement will create an axes at the top right corner of the current figure

A : subplot(2,3,3)

B : subplot(2,3,2)

C : subplot(2,3,4)

D : subplot(2,3,5)

Q.no 52. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 53. --------- function performs the custom operations for the entire dataframe.

A : function()

B : surutine()

C : rutine()

D : pipe()

Q.no 54. The --------- argument of merge function while merging two dataframes specifies which keys are to be included in the resulting dataframe.

A : right

B : on

C : sort

D : how

Q.no 55. Which of the following machine learning algorithm is used for maret basket analysis means to analyze the association of purchased items in asingle basket or single purchase.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 56. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 57. To save a figure into a file we can use ---------- method in the figure class of matplotlib.pyplot.

A : save()

B : save_fig()

C : Figure()

D : save_image()

Q.no 58. Which of the following algorithm is used in Economics, Finance, Biology etc, to model relationships between parameters of intrests.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 59. While plotting using matplotlib.pyplot A function call similar to subplot(2,3,4) is

A : subplot(234)

B : subplot(243)

C : subplot(324)

D : subplot(4)

Q.no 60. Apriori algorithm uses breadth first search and ----------structure to count candidate item sets efficiently.

A : Decision tree

B : Hash tree

C : Red-Black Tree

D : AVL Tree

Answer for Question No 1. is a

Answer for Question No 2. is a

Answer for Question No 3. is c

Answer for Question No 4. is a

Answer for Question No 5. is b

Answer for Question No 6. is d

Answer for Question No 7. is d

Answer for Question No 8. is a

Answer for Question No 9. is b

Answer for Question No 10. is d

Answer for Question No 11. is a

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is a

Answer for Question No 15. is a

Answer for Question No 16. is d

Answer for Question No 17. is c

Answer for Question No 18. is a

Answer for Question No 19. is a

Answer for Question No 20. is a

Answer for Question No 21. is d

Answer for Question No 22. is a

Answer for Question No 23. is b

Answer for Question No 24. is b

Answer for Question No 25. is d

Answer for Question No 26. is c

Answer for Question No 27. is b

Answer for Question No 28. is b

Answer for Question No 29. is b

Answer for Question No 30. is a

Answer for Question No 31. is c

Answer for Question No 32. is a

Answer for Question No 33. is b

Answer for Question No 34. is a

Answer for Question No 35. is a

Answer for Question No 36. is c

Answer for Question No 37. is a

Answer for Question No 38. is d

Answer for Question No 39. is b

Answer for Question No 40. is a

Answer for Question No 41. is a

Answer for Question No 42. is b

Answer for Question No 43. is a

Answer for Question No 44. is b

Answer for Question No 45. is a

Answer for Question No 46. is a

Answer for Question No 47. is a

Answer for Question No 48. is b

Answer for Question No 49. is a

Answer for Question No 50. is d

Answer for Question No 51. is a

Answer for Question No 52. is a

Answer for Question No 53. is d

Answer for Question No 54. is d

Answer for Question No 55. is b

Answer for Question No 56. is a

Answer for Question No 57. is b

Answer for Question No 58. is a

Answer for Question No 59. is a

Answer for Question No 60. is b

Seat No -

Total number of questions : 60

13329_DATA ANALYTICS

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. ---------- analysis estimates the relationship between single dependent variable and single independent variable

A : Simple Regression

B : Multiple regression

C : Correlation

D : Probability

Q.no 2. Find odd one out from the following :

A : KNN

B : NAïve Bayes

C : Decision Trees

D : Cluster analysis

Q.no 3. ---------- chart is a circular plot divides into sclices to show numerical proportion.

A : Bar

B : Line

C : Scatter

D : Pie

Q.no 4. ---------- type of plots show all individual data points without connected with lines.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 5. Which of the following is NOT supervised learning?

A : PCA

B : Decision Tree

C : Linear Regression

D : Naive Bayesian

Q.no 6. Probability always lies between ----- and ----

A : 0 and 1

B : -1 and +1

C : -1 and 0

D : 0 and infinite

Q.no 7. In numpy array , array indices always starts from --------

A : 1

B : -1

C : 0

D : 2

Q.no 8. To import data from excel file into a dataframe ---------- function is provided by pandas package.

A : read_csv()

B : read_file()

C : read()

D : read_excel()

Q.no 9. ---------- plot displays information as series of data points connected by straight lines.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 10. Which of the following is not a raster image file format?

A : PNG

B : JPG

C : BMP

D : PDF

Q.no 11. Naïve Bayes is a classification technique based on ----------

A : Bayes Theorem

B : Pythagorous Theorom

C : Least square method

D : mean square method

Q.no 12. ---- is an technique to learn from examples and experience, without being explicitly programmed.

A : Machine Learning

B : Software Testing

C : Computer Science

D : Data mining

Q.no 13. -------- library is built on the top of Numpy, SciPy and Matplotlib

A : Sympy

B : Scikit

C : Pandas

D : Numpy

Q.no 14. ---------- function is used to save image into an ndarray.

A : imsave()

B : imread()

C : save()

D : isave()

Q.no 15. For multidimensional visualization ---------- are used.

A : pia charts

B : Bar charts

C : Andrews curves

D : Scatter plots

Q.no 16. ---------- library from python provides efficient versions of a large number of machine learning algorithms.

A : Pandas

B : Numpy

C : Scikit-Learn

D : image

Q.no 17. In statistics, a population consists of ----------

A : All People living in a country.

B : All People living in the city.

C : All subjects or objects whose characteristics are being studied.

D : Part of whole dataset

Q.no 18. Which library from python is used for implementing machine learning algorithms?

A : Scikit-Learn

B : Pandas

C : Matplotlib

D : Numpy

Q.no 19. SQL record is an example of ----------

A : Structured data

B : Un-Structured data

C : Semi-Structured data

D : Scattered

Q.no 20. ---------- is about developing code to enable the machine to learn to perform tasks and its basic principle is the automatic modeling of underlying that have generated the collected data.

A : Data Science

B : Data Analytics

C : Data Warehousing

D : Data mining

Q.no 21. -------- is the measure of the likeihood that an event will occure in a random experiment

A : Probability

B : Correlation

C : Regression

D : Sample

Q.no 22. Entropy is a measure of the randomness in the information being processed.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 23. In head()/tail()functions of dataframe the default number of elements to display is --------

A : 3

B : 5

C : 1

D : 10

Q.no 24. In SciPy ---------- submodule is dedicated to image processing.

A : ndimage

B : ndarray

C : signal

D : io

Q.no 25. ------ module from sklearn gathers popular unsupervised clustering algorithms.

A : sklearn.covariance

B : sklearn.base

C : sklearn.neighbors

D : sklearn.cluster

Q.no 26. ---------- function used to get arrays elementwise remainder of division

A : numpy.divide(x1,x2)

B : numpy.mod(x1,x2)

C : numpy.true_divide(x1,x2)

D : numpy.reminder(x1,x2)

Q.no 27. Which of the following plots is not used for multidimensional visualization?

A : Andrrews Curves

B : Prallel Chart

C : Deviation Chart

D : Bar

Q.no 28. ---------- searches for the linear optimal separating hyperplane for separation of the data using essential training tuples called support vectors

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machines

Q.no 29. From matplotlib---------- module is used for plotting various plots.

A : Scilearn

B : Pyplot

C : Scilab

D : Matlab

Q.no 30. In ---------- the x-axes are grouped into bins and each bin will be treated as a category.

A : Bar

B : Line

C : Scatter

D : Histogram

Q.no 31. If X and Y are both independent of each other, then correlation coefficient is ---------

A : 1

B : -1

C : 0

D : 2

Q.no 32. ---------- is an indication of how often the rule has been found to be true in association rule mining.

A : Confidence

B : Support

C : Lift

D : None of These

Q.no 33. Among the following clustering algorithm types in which of the following type the notion of similarity is derived by the closeness of a data point to the centroid of the clusters.

A : Connectivity models

B : Centroid models

C : Distribution models

D : Density models

Q.no 34. The last element of ndarray is indexed by ----------

A : 0

B : -1

C : 1

D : -2

Q.no 35. ------- changes the the arrangement of items form array so that shape of array changes while maintaining the same number of dimensions.

A : numpy. Reshape()

B : numpy. Empty()

C : numpy. Flatten()

D : numpy.ravel()

Q.no 36. Identify the machine generated unstructured data.

A : Website data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 37. ---------- is unsupervised machine learning technique.

A : KNN

B : Support Vector Machines

C : Decision trees

D : Cluster analysis

Q.no 38. Support(B) =

A : (Transacions containing (B)) / (Total Transactions)

B : (Transacions containing (B)) / 100

C : (Total Transactions) / (Transacions containing (B))

D : 100/ (Transacions containing (B))

Q.no 39. ---------- is an example of semi structured data

A : XML data

B : YouTube data

C : Text File data

D : Satellite imagery data

Q.no 40. In decision trees leaf node denotes a ----------

A : class distribution

B : test on an attribute

C : outcome of the test

D : class labels

Q.no 41. Which of the following algorithm is used in Economics, Finance, Biology etc, to model relationships between parameters of intrests.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 42. In regression the dependent variable is also called as ----------

A : Regression

B : Continuous

C : Regressand

D : Independent

Q.no 43. In regression the independent variable is also called as ----------

A : Regressor

B : Continuous

C : Regressand

D : Estimated

Q.no 44. Which of the following function is not used to iterate over the rows of the DataFrame.

A : iteritems()

B : iterrows()

C : itertuples()

D : iterpanel()

Q.no 45. ---------- analysis is a set of statistical processes for estimating the relationships among dependent and independent variables.

A : Regression

B : Decision tree

C : KNN

D : None of These

Q.no 46. In unsupervised learning, scikit learn uses ---------- method to infer properties of the data.

A : extract()

B : transform()

C : infer()

D : classify()

Q.no 47. To reach to the final point and to make prediction , decision trees must be traversed from ----------

A : Top - to - bottom

B : Bottom- to - Top

C : Left- to Right

D : Right - to - Left

Q.no 48. The -- ----- is characterized by a bell shapped curve and area under curve represents probabilities

A : Normal Distribution

B : Binomial Distribution

C : Poission Distribution

D : Probability

Q.no 49. Which of the following function is used to split a figure into nrows*ncols sub-axes.

A : plot()

B : draw()

C : bar()

D : subplot()

Q.no 50. In Data science project data acquisition step involves----------

A : Acquiring data from various sources.

B : Selecting dataset

C : Data preprocessing

D : Data modeling

Q.no 51. ---------- function from scipy is used to calculate the distance between all pairs of points in a given set.

A : scipy.spatial.distance()

B : scipy.spatial.distance.measure()

C : scipy.spatial.distance.cdist()

D : distance(x1,y1)

Q.no 52. Which function returns an ndarray object that contains the numbers that are evenly spaced on a log scale.

A : numpy.logspace()

B : numpy.log()

C : numpy.fill()

D : numpy.random()

Q.no 53. In matplotlib ---------- is container class for figure instance.

A : Axes

B : Canvas

C : Figure

D : FigureCanvas

Q.no 54. ---------- machine learning algorithm used in cross marketing to work with other businesss that complement your own business but not to other competitors.

A : Decision tree

B : Association Rule Mining

C : Clustering

D : Support vector machine

Q.no 55. Select the correct statement:

A : Raw data is original source of data.

B : Preprocessed data is original source of data.

C : Raw data is the data obtained after processing steps.

D : Analysed data is original source of data.

Q.no 56. It is a measure of disorder or purity or unpredictability or uncertainty.

A : Entropy

B : Support

C : Confidence

D : lift

Q.no 57. To determine basic salary of a employee when his qualification is given is a ---------- problem

A : Correlation

B : Regression

C : Association

D : Qualitative

Q.no 58. The statement subplot(4,3,5) will divide figure into ------- and specify plotting sholud be done on plot number----------

A : 4 x 3, 5

B : 3x 4, 5

C : 3 x 5, 4

D : 5x 3, 4

Q.no 59. ---------- algorithm models a series of logical If-Then- Else decision statements, there is no underlying assumption of a linear or non-linear relationship between the input variables and response variables.

A : Regression

B : Decision Trees

C : Clustering

D : Naïve bays

Q.no 60. --------- function is used to display an image through an external viewer in scipy.

A : display()

B : imread()

C : imshow()

D : show()

Answer for Question No 1. is a

Answer for Question No 2. is d

Answer for Question No 3. is d

Answer for Question No 4. is c

Answer for Question No 5. is a

Answer for Question No 6. is a

Answer for Question No 7. is c

Answer for Question No 8. is d

Answer for Question No 9. is b

Answer for Question No 10. is d

Answer for Question No 11. is a

Answer for Question No 12. is a

Answer for Question No 13. is b

Answer for Question No 14. is a

Answer for Question No 15. is c

Answer for Question No 16. is c

Answer for Question No 17. is c

Answer for Question No 18. is a

Answer for Question No 19. is a

Answer for Question No 20. is b

Answer for Question No 21. is a

Answer for Question No 22. is a

Answer for Question No 23. is b

Answer for Question No 24. is a

Answer for Question No 25. is d

Answer for Question No 26. is b

Answer for Question No 27. is d

Answer for Question No 28. is d

Answer for Question No 29. is b

Answer for Question No 30. is d

Answer for Question No 31. is b

Answer for Question No 32. is a

Answer for Question No 33. is b

Answer for Question No 34. is b

Answer for Question No 35. is a

Answer for Question No 36. is d

Answer for Question No 37. is d

Answer for Question No 38. is a

Answer for Question No 39. is a

Answer for Question No 40. is c

Answer for Question No 41. is a

Answer for Question No 42. is c

Answer for Question No 43. is a

Answer for Question No 44. is d

Answer for Question No 45. is a

Answer for Question No 46. is b

Answer for Question No 47. is a

Answer for Question No 48. is a

Answer for Question No 49. is d

Answer for Question No 50. is a

Answer for Question No 51. is c

Answer for Question No 52. is a

Answer for Question No 53. is d

Answer for Question No 54. is b

Answer for Question No 55. is a

Answer for Question No 56. is a

Answer for Question No 57. is b

Answer for Question No 58. is a

Answer for Question No 59. is b

Answer for Question No 60. is c