



# Question Paper Preview

## Notations :

- 1.Options shown in **green** color and with  icon are correct.
- 2.Options shown in **red** color and with  icon are incorrect.

<b>Question Paper Name :</b>	M Sc Data Science 06th July 2024 Shift 3
<b>Subject Name :</b>	M.Sc. Data Science
<b>Creation Date :</b>	2024-07-06 18:29:13
<b>Duration :</b>	90
<b>Total Marks :</b>	100
<b>Display Marks:</b>	No
<b>Change Font Color :</b>	No
<b>Change Background Color :</b>	No
<b>Change Theme :</b>	No
<b>Help Button :</b>	No
<b>Show Reports :</b>	No
<b>Show Progress Bar :</b>	No

## M.Sc. Data Science

<b>Group Number :</b>	1
<b>Group Id :</b>	50131424
<b>Group Maximum Duration :</b>	0
<b>Group Minimum Duration :</b>	90
<b>Show Attended Group? :</b>	No
<b>Edit Attended Group? :</b>	No
<b>Break time :</b>	0
<b>Group Marks :</b>	100

## PART A

<b>Section Id :</b>	50131436
<b>Section Number :</b>	1
<b>Section type :</b>	Online
<b>Mandatory or Optional :</b>	Mandatory
<b>Number of Questions :</b>	100
<b>Number of Questions to be attempted :</b>	100
<b>Section Marks :</b>	100

**Maximum Instruction Time :** 0  
**Sub-Section Number :** 1  
**Sub-Section Id :** 50131441  
**Question Shuffling Allowed :** Yes

**Question Number : 1 Question Id : 5013142304 Question Type : MCQ Option Shuffling : Yes**  
**Correct Marks : 1 Wrong Marks : 0**

Processed data encompass many operations such as calculating, sorting, filtering, or summarization on unprocessed data called \_\_\_\_.

**Options :**

1. ✘ Primary data
2. ✘ Secondary data
3. ✘ Raw data
4. ✔ Preprocessing

**Question Number : 2 Question Id : 5013142305 Question Type : MCQ Option Shuffling : Yes**  
**Correct Marks : 1 Wrong Marks : 0**

The clock speed of a central processing unit measures \_\_\_\_\_

**Options :**

1. ✘ The number of units stored in cache memory
2. ✔ The processing speed of execution per second
3. ✘ The storage speed in random-access memory
4. ✘ The processing speed of storage in secondary device

**Question Number : 3 Question Id : 5013142306 Question Type : MCQ Option Shuffling : Yes**  
**Correct Marks : 1 Wrong Marks : 0**

The component in computer system responsible for computing is \_\_\_\_

**Options :**

1. ✘ Random access memory
2. ✔ Central Processing unit
3. ✘ Read only memory
4. ✘ Extendable PROM

**Question Number : 4 Question Id : 5013142307 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The correct statement about memory and storage?

**Options :**

1. ✘ Memory is temporary, storage is temporary.
2. ✔ Memory is temporary, storage is permanent.
3. ✘ Memory is permanent, storage is temporary.
4. ✘ Memory is slow, storage is fast.

**Question Number : 5 Question Id : 5013142308 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The combinational logic circuit, convert binary information from 'n' coded inputs to a maximum of '2n' unique outputs is

**Options :**

1. ✘ Multiplexer
2. ✘ Demultiplexer
3. ✘ Encoder
4. ✔ Decoder

**Question Number : 6 Question Id : 5013142309 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The octal equivalent of 1100101.001010 is \_\_\_\_\_

**Options :**

1. ✘ 624.12
2. ✔ 145.12
3. ✘ 154.12
4. ✘ 145.21

**Question Number : 7 Question Id : 5013142310 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Convert the binary equivalent 10101 to its decimal equivalent.

**Options :**

1. ✔ 21
2. ✘ 12
3. ✘ 22
4. ✘ 32

**Question Number : 8 Question Id : 5013142311 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output quality of a printer is measured in number of dots printed

**Options :**

1. ✘ per one centimetre length
2. ✘

per one inch length

3. ✓ per one square inch

4. ✗ per A4 size page

**Question Number : 9 Question Id : 5013142312 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following code is

```
set x=[1, 2, 3];
```

*If x is a list, print length of x, else print "Not a list"*

**Options :**

1. ✓ 3

2. ✗ 0

3. ✗ error

4. ✗ Not a list

**Question Number : 10 Question Id : 5013142313 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The model best suited when the customer/user software requirements are not clear / decided, then the user interface is to look like

**Options :**

1. ✗ Classical waterfall model

2. ✓ Prototyping model

3. ✗ Evolutionary model

4. ✗ Logit model

**Question Number : 11 Question Id : 5013142314 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Not a characteristic among the following for a good software design is

**Options :**

1. ✘ Correctness
2. ✘ Understandability
3. ✔ Complex design
4. ✘ Efficiency

**Question Number : 12 Question Id : 5013142315 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Software designers free to make any alterations, corrections and modifications in the phase \_\_\_\_ .

**Options :**

1. ✔ Preliminary design phase
2. ✘ Detailed design phase
3. ✘ End phase
4. ✘ Implementation phase

**Question Number : 13 Question Id : 5013142316 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_ version of Python removed the print statement.

**Options :**

1. ✘ 1.x

2. ✘ 2.x

3. ✔ 3.x

4. ✘ 4.x

**Question Number : 14 Question Id : 5013142317 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The number of Lagrange multipliers required for estimating support vector machine training parameters depends on

**Options :**

1. ✘ The dimension of the data

2. ✘ The number of classes

3. ✔ The total number of support vectors from training samples

4. ✘ Number of variables under study

**Question Number : 15 Question Id : 5013142318 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following python code is

```
print("Python", "Programming", sep="-")
```

**Options :**

1. ✘ PythonProgramming

2. ✘ Python Programming

3. ✘ Python,Programming

4. ✓ Python-Programming

**Question Number : 16 Question Id : 5013142319 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In the *elif* and *else* blocks are missing in an if statement the output is

**Options :**

1. ✘ Syntax error
2. ✘ The program will stop
3. ✓ Nothing, it's optional
4. ✘ Runtime error

**Question Number : 17 Question Id : 5013142320 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Given a class Animal and a subclass Dog, the method demonstrates polymorphism is

**Options :**

1. ✓ Animal.speak() and Dog.speak() have different implementations
2. ✘ Dog has the same attributes as Animal
3. ✘ Dog uses a method from Animal without changing it
4. ✘ Animal has a method not present in Dog

**Question Number : 18 Question Id : 5013142321 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**



`list.index(x[, start[, end]])` is used to

**Options :**

1. ✘ Return zero-based index in the list
2. ✘ Raises a ValueError if there is no such item
3. ✔ Both Return zero-based index in the list and Raises a ValueError if there is no such item
4. ✘ Error

**Question Number : 19 Question Id : 5013142322 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following code snippet is

```
x = 5y = 2print(x ** y)
```

**Options :**

1. ✘ 10
2. ✘ 7
3. ✔ 25
4. ✘ 3

**Question Number : 20 Question Id : 5013142323 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

`strip()` method in Python

**Options :**

1. ✔ Removes all leading and trailing spaces from a string
2. ✘ Splits a string into a list of substrings

3. ✘ Joins multiple strings into one
4. ✘ Replaces occurrences of a substring within a string

**Question Number : 21 Question Id : 5013142324 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

What does the `__init__` method do in Python

**Options :**

1. ✘ It initializes all the variables in a class.
2. ✔ It is a constructor method that is automatically called when a new instance of a class is created.
3. ✘ It is used to delete an instance of a class.
4. ✘ It is a reserved keyword and cannot be used as a method name.

**Question Number : 22 Question Id : 5013142325 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The `seek()` method is used to \_\_\_\_.

**Options :**

1. ✘ Save the file in secondary storage
2. ✔ Position the file object at a particular position in a file
3. ✘ Delete the file form secondary storage
4. ✘ Hiding the file / object stored

**Question Number : 23 Question Id : 5013142326 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The module Pickle is used to \_\_\_\_.

**Options :**

1. ✓ Serializing / De-serializing Python object structure
2. ✗ Pick and remove the object
3. ✗ Convert
4. ✗ Save the object

**Question Number : 24 Question Id : 5013142327 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Assume a dataset containing information about 200 individuals. One hundred of these individuals have purchased life insurance. A supervised data mining session has discovered the following rule

IF age < 30 & credit card insurance = yes THEN life insurance = yes

Rule Accuracy: 70%

Rule Coverage: 63%

How many individuals in the class life insurance= no have credit card insurance and are less than 30 years old?

**Options :**

1. ✗ 63
2. ✗ 70
3. ✗ 30
4. ✓ 18

**Question Number : 25 Question Id : 5013142328 Question Type : MCQ Option Shuffling : Yes**

**Correct Marks : 1 Wrong Marks : 0**

The error in the following code is

```
try:  
    x = 1 / "0"  
except ValueError:  
    print("Type error occurred")
```

**Options :**

1. ✘ Wrong exception type is caught
2. ✘ Should be *ZeroDivisionError*
3. ✔ It will raise a *TypeError*
4. ✘ No error

**Question Number : 26 Question Id : 5013142329 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The correct example on the usage of JSON object with name, age, and city is

**Options :**

1. ✔ `{ "record": {"name": "Alex", "age": 21, "city": "London"} }`
2. ✘ `{ "record"= {"name": "Alex", "age": 21, "city": "London"} }`
3. ✘ `{ "record": {"name": Alex, "age": 21, "city": London} }`
4. ✘ `{ record : {"name": "Alex", "age": 21, "city": "London"} }`

**Question Number : 27 Question Id : 5013142330 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The function used to convert a JavaScript object into a string is

**Options :**

1. ✘ *JSON.string()*
2. ✔ *JSON.stringify()*
3. ✘ *JSON.ToString()*
4. ✘ *JSONString.fy()*

**Question Number : 28 Question Id : 5013142331 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Python module is a

**Options :**

1. ✘ built-in function
2. ✔ collection of Python functions and global variables
3. ✘ type of Python data structure
4. ✘ programming language

**Question Number : 29 Question Id : 5013142332 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The module used in Python for sending emails is

**Options :**

1. ✘ *email*
2. ✘ *sendmail*
3. ✘ *mail*

4. ✓ `smtplib`

**Question Number : 30 Question Id : 5013142333 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

To draw a plot without a line, correct code is

**Options :**

1. ✗ `plt.plot(xpoints, ypoints)`

2. ✗ `plt.plot(xpoints, ypoints, 0)`

3. ✗ `plt.plot(xpoints, ypoints, False)`

4. ✓ `plt.plot(xpoints, ypoints, 'o')`

**Question Number : 31 Question Id : 5013142334 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The parameter used to define the position/alignment of the plot title in the `plt.title()` function is

**Options :**

1. ✗ `pos`

2. ✗ `align`

3. ✓ `loc`

4. ✗ `justify`

**Question Number : 32 Question Id : 5013142335 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

**Numpy** developed by

**Options :**

1. ✘ Guido van Rossum
2. ✔ Travis Oliphant
3. ✘ Wes McKinney
4. ✘ Jim Hugunin

**Question Number : 33 Question Id : 5013142336 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

*ndarray* is created using \_\_\_\_\_

**Options :**

1. ✘ `numpy.array(object, dtype = None, copy = True, subok = False, ndmin = 0)`  
`numpy.array(object, dtype = None, copy = True, order = None, subok = False, ndmin = 0)`
2. ✔ `numpy.array(object, dtype = None, copy = True, order = None, subok = False, ndmin = 0)`
3. ✘ `numpy_array(object, dtype = None, copy = True, order = None, subok = False, ndmin = 0)`
4. ✘ `numpy.array(object, dtype = None, copy = True, order = None, ndmin = 0)`

**Question Number : 34 Question Id : 5013142337 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Wrong statement in the following is \_\_\_\_\_

**Options :**

1. ✔ *ndarray* is also known as the axis array.

2. ✘ ndarray.dataitemSize is the buffer containing the actual elements of the array.
3. ✘ NumPy main object is the homogeneous multidimensional array.
4. ✘ In Numpy, dimensions are called axes

**Question Number : 35 Question Id : 5013142338 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following code is \_\_\_\_

```
import numpy as np
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
c = np.stack((a, b))
print(c)
```

**Options :**

1. ✘ `[[1, 2, 3], [4, 5, 6]]`
2. ✔ `[[1, 4], [2, 5], [3, 6]]`
3. ✘ `[1, 2, 3, 4, 5, 6]`
4. ✘ Error

**Question Number : 36 Question Id : 5013142339 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**



The output of the following code is

```
import numpy as np
a = np.array([[1, 2], [3, 4]])
b = np.array([[5, 6], [7, 8]])
c = np.dot(a, b)
print(c)
```

Options :

1. ✓  `[[19, 22], [43, 50]]`
2. ✗  `[[5, 6], [7, 8], [1, 2], [3, 4]]`
3. ✗  `[[1, 5], [2, 6], [3, 7], [4, 8]]`
4. ✗  Error

Question Number : 37 Question Id : 5013142340 Question Type : MCQ Option Shuffling : Yes  
Correct Marks : 1 Wrong Marks : 0

The distance of a feature vector  $[2, 3, -2]$  from separating plane  $x_1 + 2x_2 + 2x_3 + 5 = 0$  is given by

Options :

1. ✗  5
2. ✗   $3/13$
3. ✓  3
4. ✗  13

Question Number : 38 Question Id : 5013142341 Question Type : MCQ Option Shuffling : Yes  
Correct Marks : 1 Wrong Marks : 0

The response in the execution of following code is \_\_\_\_

```
import pandas as pd  
  
s =pd.Series([1,2,3,4,5],index= ['a','b','c','d','e'])  
  
print(s['f'])
```

**Options :**

1. ✓  KeyError
2. ✗  IndexError
3. ✗  ValueError
4. ✗  a=1,b=2,c=3,d=4

**Question Number : 39 Question Id : 5013142342 Question Type : MCQ Option Shuffling : Yes**

**Correct Marks : 1 Wrong Marks : 0**

correct syntax for data frame in pandas is \_\_\_\_

**Options :**

1. ✓  `pandas.DataFrame(data, index, dtype, copy)`
2. ✗  `pandas.DataFrame( data, index, columns, dtype, copy)`
3. ✗  `pandas.DataFrame(data, index, dtype, dcopy)`
4. ✗  `pandas.DataFrame( data, index, rows, dtype, copy)`

**Question Number : 40 Question Id : 5013142343 Question Type : MCQ Option Shuffling : Yes**

**Correct Marks : 1 Wrong Marks : 0**

The wrong statement(s) in the following

- a) *iteritems()* returns each column's value in form of series object.
- b) *tail()* returns any number of bottom rows by specifying values of number's argument.

**Options :**

- 1. ✘ (a)
- 2. ✘ (b)
- 3. ✘ (a) & (b)
- 4. ✔ (a) & (b) both correct

**Question Number : 41 Question Id : 5013142344 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The wrong statement(s) in the following

- a) *writelines()* method is used to write multiple strings to a file.
- b) Indexing in series is similar to that for NumPy arrays

**Options :**

- 1. ✘ (a)
- 2. ✘ (b)
- 3. ✘ (a) & (b)
- 4. ✔ (a) & (b) both correct

**Question Number : 42 Question Id : 5013142345 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

To provide easy plotting of data, project build on top of 'pandas' and 'matplotlib' is

**Options :**

1. ✘ Yhat
2. ✔ Seaborn
3. ✘ Vincent
4. ✘ Pychart

**Question Number : 43 Question Id : 5013142346 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The purpose of the *cursor()* method in Python's database interaction is to

**Options :**

1. ✘ create a new database
2. ✔ execute SQL queries
3. ✘ close the database connection
4. ✘ fetch all records from a table

**Question Number : 44 Question Id : 5013142347 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In Python's database interaction, *fetchone()* method does

**Options :**

1. ✘ Fetches all rows from the result set
2. ✘ Fetches the next row of a query result set

3. ✓ Fetches the first row of a query result set
4. ✘ Fetches a specific row based on index

**Question Number : 45 Question Id : 5013142348 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Correct statement(s) in the following on *MongoDB*?

- a) *MongoDB* is a cross-platform database
- b) *MongoDB* is document-oriented database
- c) *MongoDB* is high performance database

**Options :**

1. ✘ (a) & (b)
2. ✘ (b) & (c)
3. ✘ (a) & (c)
4. ✓ (a), (b) & (c)

**Question Number : 46 Question Id : 5013142349 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Correct statement(s) in the following is / are

- a) The collections screen displays collection name for each collection in the selected database.
- b) The collections screen displays number of documents in the collection for each collection in the selected database.

**Options :**

1. ✘ (a)

2. ✘ (b)

3. ✔ (a) & (b)

4. ✘ (a) & (b) both wrong

**Question Number : 47 Question Id : 5013142350 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

*re.search()* function in Python used to

**Options :**

1. ✔ Search for a pattern within a string

2. ✘ Sort a list of strings

3. ✘ Convert a string to lowercase

4. ✘ Compare two strings

**Question Number : 48 Question Id : 5013142351 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following Python code is

```
sentence = 'we are humans'
```

```
matched = re.match(r'(.*) (.*) (.*)', sentence)
```

```
print(matched.groups())
```

**Options :**

1. ✔ ('we', 'are', 'humans')

2. ✘ (we, are, humans)

3. ✘ ('we', 'humans')

'we are humans'

4. ✘

**Question Number : 49 Question Id : 5013142352 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The output of the following Python code is

```
sentence = 'horses are fast'  
regex = re.compile('(P<animal>\w+) (P<verb>\w+) (P<adjective>\w+)')  
matched = re.search(regex, sentence)  
print(matched.groups())
```

**Options :**

1. ✘ {'animal': 'horses', 'verb': 'are', 'adjective': 'fast'}

2. ✔ ('horses', 'are', 'fast')

3. ✘ 'horses are fast'

4. ✘ 'are'

**Question Number : 50 Question Id : 5013142353 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The *re.sub()* function in Python

**Options :**

1. ✘ Searches for a pattern within a string and returns the first match

2. ✘ Finds all matches of a pattern within a string and returns them as a list

3. ✔ Replaces all occurrences of a pattern with a specified string

Splits a string into a list based on a pattern

4. ✖

**Question Number : 51 Question Id : 5013142354 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Correct statement(s) among the following on Natural language processing (NLP) and Machine learning (ML) is

- a) NLP is a type specific subfield of artificial intelligence that focuses on processing and understanding human language whereas ML techniques are not applicable.
- b) ML is a set of algorithms used to train the numeric data whereas NLP uses other than ML algorithms
- c) NLP is focused on language-specific tasks, while ML is more general.
- d) No difference between NLP and ML

**Options :**

1. ✖ (a)

2. ✖ (b)

3. ✔ (c)

4. ✖ (d)

**Question Number : 52 Question Id : 5013142355 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

A rule-based approach to natural language processing is

**Options :**

1. ✖ Using machine learning algorithms to analyze text

2. ✔ Creating a set of if-then rules to analyze text

3. ✖ Using neural networks to translate text



Analysing text using genetic algorithms

4. ✘

**Question Number : 53 Question Id : 5013142356 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Morphological segmentation is

**Options :**

1. ✘ Used to do discourse analysis
2. ✘ an extension of propositional logic
3. ✔ Separate words into individual morphemes and identify the class of the morphemes
4. ✘ tokenization of complex statements

**Question Number : 54 Question Id : 5013142357 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Many words have more than one meaning; To select the meaning which makes the most sense in context, can be resolved using

**Options :**

1. ✘ Fuzzy logic
2. ✔ Word sense disambiguation
3. ✘ Shallow semantic analysis
4. ✘ Kalman word filter controllers

**Question Number : 55 Question Id : 5013142358 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following uses probabilistic approach to speech recognition

**Options :**

1. ✓ Hidden markov model
2. ✗ Fuzzy rule-based system
3. ✗ Recurrent neural network
4. ✗ Convolutional neural network

**Question Number : 56 Question Id : 5013142359 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

A common evaluation metric used for machine translation is

**Options :**

1. ✗ Precision
2. ✗ Recall
3. ✗ F1 score
4. ✓ BLEU score

**Question Number : 57 Question Id : 5013142360 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following is an example of a natural language processing application in the healthcare industry?

**Options :**

1. ✗ Speech recognition for virtual assistants

- Sentiment analysis for social media
2. ✘
- Machine translation for international business
3. ✘
- Clinical text mining for electronic health records
4. ✔

**Question Number : 58 Question Id : 5013142361 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The simplest NoSQL databases is

**Options :**

1. ✔ Key-value stores
2. ✘ Document Databases
3. ✘ Wide-column
4. ✘ Wide-row

**Question Number : 59 Question Id : 5013142362 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

NoSQL is a database

**Options :**

1. ✘ that is an enhanced form of RDBMS.
2. ✘ that is built with enhancements to DBMS.
3. ✔ that is built on ways and means other than tables and columns.
4. ✘ that does not support an enhanced form of RDBMS

**Question Number : 60 Question Id : 5013142363 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

A set of data that converts into another set of data, where individual elements are broken down into tuples (key/value pairs) is called

**Options :**

1. ✓ Map
2. ✗ Reduce
3. ✗ Transformation
4. ✗ Node

**Question Number : 61 Question Id : 5013142364 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Point out the correct statement.

- a) MapReduce tries to place the data and the compute as close as possible.
- b) Map Task in MapReduce is performed using the Mapper() function.
- c) Reduce Task in MapReduce is performed using the Map() function

**Options :**

1. ✓ (a)
2. ✗ (b)
3. ✗ (c)
4. ✗ (a) & (b)

**Question Number : 62 Question Id : 5013142365 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following is not accurately describe Hadoop\_\_\_\_\_

**Options :**

1. ✘ Open-source
2. ✔ Real-time
3. ✘ Java-based
4. ✘ Distributed computing approach

**Question Number : 63 Question Id : 5013142366 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information is

**Options :**

1. ✔ Data Mining
2. ✘ Data Warehousing
3. ✘ Data Integration
4. ✘ Data Processing

**Question Number : 64 Question Id : 5013142367 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following is not a common use case for Big Data analytics?

**Options :**

1. ✘ Fraud Detection
2. ✘ Customer Segmentation
3. ✘

## Social Media Analysis

### Inventory Management

4. ✓

**Question Number : 65 Question Id : 5013142368 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The basic measure used for a model prediction change based on interval scale input is

**Options :**

1. ✗ Bias

2. ✓ Variance

3. ✗ Precision

4. ✗ Recall

**Question Number : 66 Question Id : 5013142369 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Hadoop framework is implemented in Java, MapReduce applications need not be written in

**Options :**

1. ✗ C

2. ✗ C#

3. ✓ Java

4. ✗ Visual C++

**Question Number : 67 Question Id : 5013142370 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The Hadoop language developed by

**Options :**

1. ✘ Bell Labs
2. ✘ Sun Microsystems
3. ✔ Apache Software Foundation
4. ✘ Hadoop Software Foundation

**Question Number : 68 Question Id : 5013142371 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The Hadoop list includes the HBase database, the Apache Mahout \_\_\_ system, and matrix operations.

**Options :**

1. ✘ Artificial intelligence
2. ✔ Machine learning
3. ✘ Pattern recognition.
4. ✘ Statistical classification.

**Question Number : 69 Question Id : 5013142372 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following model(s) can perform tweet classification with regards to context mentioned data of tweets

**Options :**

1. ✘

Naïve Bayes

2. ✘ Support Vector Machine
3. ✘ Multiple Linear Regression
4. ✔ Unsupervised Learning Technique

**Question Number : 70 Question Id : 5013142373 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Let a document term matrix of the data is created, treating every tweet as one document. in regards to the same, which of the following is correct?

- a) Removal of stop words from the data will affect the dimensionality of data.
- b) Normalization of words in the data will reduce the dimensionality of data.
- c) Converting all the words in lowercase will not affect the dimensionality of the data.

**Options :**

1. ✘ (a)
2. ✘ (b)
3. ✘ (c)
4. ✔ (a), (b)

**Question Number : 71 Question Id : 5013142374 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following is not a characteristic of deep learning algorithms?

**Options :**

1. ✘ Ability to automatically learn features



2. ✓ Dependence on human-designed features
3. ✗ Scalability with large datasets
4. ✗ Utilization of neural network architectures

**Question Number : 72 Question Id : 5013142375 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The primary purpose of an activation function in a neural network is

**Options :**

1. ✓ To introduce non-linearity into the model.
2. ✗ To initialize the weights.
3. ✗ To optimize the loss function.
4. ✗ To scale the input data.

**Question Number : 73 Question Id : 5013142376 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Let  $V = (2\ 4\ 6)$  be a vector then  $L_2$  length of the vector is

**Options :**

1. ✗ 12
2. ✗ 28
3. ✓ 7.483
4. ✗

**Question Number : 74 Question Id : 5013142377 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In gradient-based optimization, what does the gradient of the loss function with respect to the model parameters represent?

**Options :**

1. ✓ The direction and rate of change to minimize the loss
2. ✗ The best parameters for the model
3. ✗ The loss value
4. ✗ The learning rate

**Question Number : 75 Question Id : 5013142378 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The best statement that describes Keras among the following is

**Options :**

1. ✗ A low-level deep learning framework requiring detailed tensor manipulations.
2. ✓ A high-level neural networks API, written in Python and capable of running on top of Tensor Flow, CNTK, or Theano.
3. ✗ An open-source machine learning library primarily for data preprocessing.
4. ✗ A purely backend deep learning library optimized for running on GPUs only

**Question Number : 76 Question Id : 5013142379 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In linear regression, what is the effect of multicollinearity among predictor variables on the model coefficients?

**Options :**

1. ✓ It causes coefficients to increase in magnitude
2. ✗ It causes coefficients to decrease in magnitude
3. ✗ It introduces bias in coefficient estimates
4. ✗ It has no effect on coefficient estimates

**Question Number : 77 Question Id : 5013142380 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In using gradient descent-based optimization algorithms, the feature scaling is important because it

**Options :**

1. ✗ reduces the dimensionality of the data
2. ✗ speeds up the convergence of the algorithm
3. ✗ improves the interpretability of the model
4. ✓ prevents numerical instability.

**Question Number : 78 Question Id : 5013142381 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

A logistic regression model is trained for binary classified data and found that the probabilities, for class 0 and class 1 are 0.3 and 0.7 respectively. If a decision threshold value is 0.5, then the predicted class is

**Options :**

1. ✘ Class 0
2. ✔ Class 1
3. ✘ Not enough information to determine
4. ✘ Class label depends on the data set and its distribution

**Question Number : 79 Question Id : 5013142382 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Let A (1, 2), B (2, 3), C (3, 4), and D (4, 5) be the two-dimensional data set. Using K-nearest neighbour classifier with  $k=3$ , and based on majority voting predicted class for data point E (3, 3) is

**Options :**

1. ✘ Class A
2. ✔ Class B
3. ✘ Class C
4. ✘ Class D

**Question Number : 80 Question Id : 5013142383 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In a dataset, 80% of emails contain the word "discount" are spam. If 10% of all emails are spam and an email contains the word discount, what's the probability that it is spam?

**Options :**

1. ✘ 0.6923
2. ✔ 0.3077
3. ✘

0.421

4. ✘ 0.8

**Question Number : 81 Question Id : 5013142384 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which clustering algorithm is robust to outliers in the data?

**Options :**

1. ✘ K-means clustering

2. ✔ DBSCAN

3. ✘ Hierarchical clustering

4. ✘ Gaussian mixture models

**Question Number : 82 Question Id : 5013142385 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The primary advantage of using Gaussian mixture models over k-means clustering

**Options :**

1. ✘ is simpler to implement

2. ✘ can capture non-linear relationships

3. ✔ provides probabilistic cluster assignments

4. ✘ is more computationally efficient

**Question Number : 83 Question Id : 5013142386 Question Type : MCQ Option Shuffling : Yes**

**Correct Marks : 1 Wrong Marks : 0**

Let there be a trained model in Keras with size of 50 MB. If `model.save('model.h5')` is used to save it, the space of saved model is approximately

**Options :**

1. ✘ 25 MB
2. ✔ 50 MB
3. ✘ 75 MB
4. ✘ 100 MB

**Question Number : 84 Question Id : 5013142387 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Identify the correct statements

- a) Data Normalization is a technique that is used to convert a numeric variable into a specified range such as  $[0,1]$ .
- b) Z-Score Normalization is a technique that is used to convert a variable that has a 0 mean and 1 standard deviation using Mean and standard deviation of that variable.
- c) Log transformation is used to convert it into a normal distribution when the data does not conform to the normal distribution and has a skewed distribution.
- d) Square root transformation is used to convert to features that are having only positive values and replaced the values by its square root values for the normality.

**Options :**

1. ✘ (a) & (b)
2. ✘ (b) & (c)
3. ✔ (a), (b) & (c)

4. ✘ (a), (b), (c) & (d)

**Question Number : 85 Question Id : 5013142388 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The metric commonly used for classification tasks when class imbalance is present is

**Options :**

1. ✘ Mean squared error

2. ✘ Accuracy

3. ✔ F<sub>1</sub>-score

4. ✘ R<sup>2</sup>-score

**Question Number : 86 Question Id : 5013142389 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The type of machine learning algorithm aims to mimic the process of human learning?

**Options :**

1. ✘ Supervised learning

2. ✘ Unsupervised learning

3. ✔ Reinforcement learning

4. ✘ Deep learning

**Question Number : 87 Question Id : 5013142390 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which technique is used to reduce the dimensionality of data while preserving as much information as possible?

**Options :**

1. ✘ Clustering
2. ✔ Feature extraction
3. ✘ Feature selection
4. ✘ Regularization

**Question Number : 88 Question Id : 5013142391 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which algorithm is used for hierarchical clustering?

**Options :**

1. ✘ K-Means clustering
2. ✔ Agglomerative clustering
3. ✘ DBSCAN
4. ✘ Principal Component Analysis

**Question Number : 89 Question Id : 5013142392 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which type of learning is characterized by an agent learning through interactions with an environment and receiving rewards?

**Options :**

1. ✘ Supervised learning
2. ✘ Unsupervised learning



3. ✓ Reinforcement learning

4. ✗ Semi-supervised learning

**Question Number : 90 Question Id : 5013142393 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The primary purpose of a decision tree's leaf nodes is to

**Options :**

1. ✓ make predictions

2. ✗ split the data

3. ✗ represent features

4. ✗ store feature values

**Question Number : 91 Question Id : 5013142394 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

If the error information can't be used to improve the neural network behavior is called

**Options :**

1. ✗ Supervised

2. ✗ Unsupervised

3. ✓ Reinforcement

4. ✗ Kalman filter controller

**Question Number : 92 Question Id : 5013142395 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Gradient descent is a direction towards optimal point

Options :

1. ✘ neutral
2. ✔ negative
3. ✘ positive
4. ✘ not depends on the direction

Question Number : 93 Question Id : 5013142396 Question Type : MCQ Option Shuffling : Yes  
Correct Marks : 1 Wrong Marks : 0

Which of the following is an unsupervised learning algorithm?

Options :

1. ✘ Perceptron
2. ✘ Backpropagation
3. ✘ Self-Organising Map
4. ✔ Hebbian

Question Number : 94 Question Id : 5013142397 Question Type : MCQ Option Shuffling : Yes  
Correct Marks : 1 Wrong Marks : 0

Among the following which is a log linear model?

Options :

1. ✘  $Y=ax+b$
2. ✔  $Y= ax^b$
3. ✘

$$Y = ax^d + bx + c$$

4. ✘  $Y = \exp(a \cdot \log(x^b) + c)$

**Question Number : 95 Question Id : 5013142398 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The Robust regression methods are applied when errors

**Options :**

1. ✘ follows Normal
2. ✔ follows non-normal
3. ✘ mean square is large
4. ✘ are independent

**Question Number : 96 Question Id : 5013142399 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Let a response variable is measured on interval scale and explanatory variable is measured on categorical the model applicable is

**Options :**

1. ✘ Simple logistic regression model
2. ✘ Simple linear regression model
3. ✔ Probit Regression model
4. ✘ Orthogonal Polynomial Regression model

**Question Number : 97 Question Id : 5013142400 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

For a simple linear regression model to test the hypothesis that  $H_0: \beta_1 = 0$  versus  $H_1: \beta_1 \neq 0$  the test statistic used is

**Options :**

1. ✘ t-test
2. ✔ F- test
3. ✘  $\chi^2$ - test
4. ✘ Z-test

**Question Number : 98 Question Id : 5013142401 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

In regression analysis, the  $R^2$  statistic is in usual notation

**Options :**

1. ✘  $S_{xx}^2 / S_{yy}^2$
2. ✘  $S_{yy}^2 / S_{xx}^2$
3. ✔  $SS_{Reg} / SS_{Tot}$
4. ✘  $(SS_{Reg}/k) / (SS_{Tot}/n-k)$

**Question Number : 99 Question Id : 5013142402 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

The difference between standard and un-standard multiple linear regression models is

**Options :**

1. ✘ Standard model is evaluated for standard normal variates data whereas un-standard model is evaluated for non-normal data

2. ✓ In standard model constant is zero whereas for un-standard model it is not zero
3. ✗ Both models are same with differing variances in the estimated response
4. ✗ Both are evaluated for non-normal

**Question Number : 100 Question Id : 5013142403 Question Type : MCQ Option Shuffling : Yes Correct Marks : 1 Wrong Marks : 0**

Which of the following is a growth curve

**Options :**

1. ✓ Logistic curve
2. ✗ Probit curve
3. ✗ Polynomial curve
4. ✗ Multiple linear curve