

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

MCA

SEM: IV - THEORY EXAMINATION (20..... - 20.....)

Subject: Business Data Analytics

Time: 3 Hours

Max. Marks: 100

## General Instructions:

IMP: Verify that you have received the question paper with the correct course, code, branch etc.

1. This Question paper comprises of three Sections -A, B, &amp; C. It consists of Multiple Choice Questions (MCQ's) &amp; Subjective type questions.

2. Maximum marks for each question are indicated on right -hand side of each question.

3. Illustrate your answers with neat sketches wherever necessary.

4. Assume suitable data if necessary.

5. Preferably, write the answers in sequential order.

6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

**SECTION-A**

20

1. Attempt all parts:-

1-a. Which of the following technologies marked the 'Big Data Era'? (CO1, K1)

1

- (a) SPSS and SAS
- (b) SQL and relational databases
- (c) Hadoop and Spark
- (d) Microsoft Excel

1-b. What enabled faster data processing during the mid-20th century? (CO1, K1)

1

- (a) Internet
- (b) Computers
- (c) Smartphones
- (d) Cloud computing

1-c. What does the mean of a dataset represent?(CO2, K1)

1

- (a) Middle value
- (b) Most frequently occurring value
- (c) Average value
- (d) Spread of the data

1-d. Which function is used to calculate median in Python? (CO2, K1)

1

- (a) np.mean()
- (b) np.median()
- (c) np.mode()

- (d) `np.var()`
- 1-e. Which function removes duplicate rows from a DataFrame? (CO3, K1) 1
- (a) `remove_duplicates()`
  - (b) `drop_duplicates()`
  - (c) `delete_duplicates()`
  - (d) `clear_duplicates()`
- 1-f. What method is used to combine DataFrames horizontally in pandas?(CO3, K1) 1
- (a) `concat()`
  - (b) `merge()`
  - (c) `append()`
  - (d) `join()`
- 1-g. What is a key element of Machine Learning?(CO4, K1) 1
- (a) Big Data visualization
  - (b) Model evaluation
  - (c) Page layout
  - (d) Template design
- 1-h. Supervised learning requires: (CO4, K1) 1
- (a) Manual rule writing
  - (b) Data with output labels
  - (c) Unstructured data
  - (d) Encrypted data
- 1-i. Which metric is not used to evaluate classification models? (CO5, K1) 1
- (a) Accuracy
  - (b) Recall
  - (c) Mean Squared Error
  - (d) F1 Score
- 1-j. Which logistic regression type handles more than two categories?(CO5, K1) 1
- (a) Binary
  - (b) Ordinal
  - (c) Multinomial
  - (d) Linear

2. Attempt all parts:-

- 2.a. Define Big Data characteristics. (CO1, K1) 2
- 2.b. Explain measures of central tendency.(CO2, K2) 2
- 2.c. Name two types of data processing. (CO3, K1) 2
- 2.d. Name 2 types of Machine Learning.(CO4, K1) 2
- 2.e. List types of logistic regression. (CO5, K1) 2

## **SECTION-B**

30

3. Answer any five of the following:-

- 3-a. Explain the benefits of predictive modelling.(CO1, K2) 6
- 3-b. Summarize the use of sentiment analysis in text mining.(CO1, K4) 6
- 3-c. Summarize the concept of permutation testing. (CO2, K4) 6
- 3-d. Outline the key benefits of permutation tests. (CO2, K2) 6
- 3.e. How does one-hot encoding handle categorical data? (CO3, K2) 6
- 3.f. Discuss the role of multicollinearity. (CO4,K2) 6
- 3.g. Write impact of multicollinearity. (CO5, K2) 6

## **SECTION-C**

50

4. Answer any one of the following:-

- 4-a. Discuss the characteristics of Big Data and its significance in analytics. (CO1, K2) 10
- 4-b. Provide a detailed overview of SQL's role in managing relational databases.(CO1, K3) 10

5. Answer any one of the following:-

- 5-a. Explain standard deviation and variance with a dataset example.(CO2, K2) 10
- 5-b. Illustrate skewness with a practical example and its significance.(CO2, K2) 10

6. Answer any one of the following:-

- 6-a. How do clustering techniques (e.g., DBSCAN) identify outliers? (CO3, K2) 10
- 6-b. Discuss the impact of outliers on statistical models.(CO3, K2) 10

7. Answer any one of the following:-

- 7-a. Compare supervised, unsupervised, and reinforcement learning with examples. (CO4, K4) 10
- 7-b. Describe the working of supervised learning using a real-world example.(CO4, K2) 10

8. Answer any one of the following:-

- 8-a. Discuss steps in logistic regression model building.(CO5, K2) 10
- 8-b. Differentiate between logistic and linear regression.(CO5, K4) 10