

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

MCA

SEM: IV - THEORY EXAMINATION (2025 - 2026)

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, & C**. It consists of Multiple Choice Questions (MCQ's) & 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:-

- (1.1) Early statistical software like SPSS and SAS emerged during:(CO1,K1) 1
- (a) The pre-digital era
- (b) The mid-20th century
- (c) The 1990s
- (d) The Big Data Era
- (1.2) A/B testing is used for:(CO1,K2) 1
- (a) Testing database queries
- (b) Comparing webpage versions
- (c) Automating analysis
- (d) Real-time analytics
- (1.3) Which function is used to calculate median in Python? (CO2, K1) 1
- (a) np.mean()
- (b) np.median()
- (c) np.mode()
- (d) np.var()
- (1.4) What is inferential statistics used for?(CO2,K2) 1
- (a) Summarizing past data
- (b) Making predictions based on sample data
- (c) Creating visualizations
- (d) Measuring data dispersion
- (1.5) What method is used to filter rows in pandas based on a condition? (CO3, K1) 1

- (a) filter()
 - (b) select()
 - (c) query()
 - (d) where()
- (1.6) Which technique is NOT used for missing value treatment?(CO3,K2) 1
- (a) Deletion
 - (b) Imputation
 - (c) Normalization
 - (d) Forward fill
- (1.7) What does Multiple Linear Regression involve?(CO4,K2) 1
- (a) Multiple output variables
 - (b) One input variable
 - (c) Multiple input variables
 - (d) No variables
- (1.8) Polynomial Regression is an extension of:(CO4,K2) 1
- (a) KNN
 - (b) Linear Regression
 - (c) Clustering
 - (d) SVM
- (1.9) What is a key feature of Random Forest?(CO5,K2) 1
- (a) Single tree
 - (b) Unsupervised
 - (c) Ensemble of trees
 - (d) Linear separation
- (1.10) Which algorithm works well with text data?(CO5,K2) 1
- (a) Naive Bayes
 - (b) Linear Regression
 - (c) KNN
 - (d) Random Forest

2. Attempt all parts:-

- (2.1) Define Big Data characteristics. (CO1, K1) 2
- (2.2) Explain various measures of dispersion.(CO2,K2) 2
- (2.3) What is data processing? (CO3, K2) 2
- (2.4) Describe reinforcement learning.(CO4,K2) 2
- (2.5) Define feature scaling.(CO5,K2) 2

SECTION-B

30

Answer any one of the following:-

- (3.1) Summarize the use of sentiment analysis in text mining.(CO1, K4) 6
- (3.2) Highlight differences between structured and unstructured data in detail with example.(CO1,K4) 6

Answer any one of the following:-

- (3.3) Illustrate the relationship between standard deviation and variance.(CO2,K2) 6
- (3.4) Specify the role of normal distribution in hypothesis testing.(CO2,K2) 6

Answer any one of the following:-

- (3.5) Elaborate the concept of discretization, and how does it work?(CO3,K2) 6
- (3.6) Discuss the role of data transformation in time-series analysis.(CO3,K2) 6

Answer any one of the following:-

- (3.7) Briefly explain various applications of Machine Learning in field of education.(CO4,K2) 6
- (3.8) Explain the role of labeled data in supervised learning[K2,CO4] 6

Answer any one of the following:-

- (3.9) Discuss use of NLP in classification.(CO5,K2) 6
- (3.10) Describe memory feature of SVM.(CO5,K2) 6

SECTION-C

50

4. Answer any one of the following:-

- (4.1) Discuss the impact of real-time analytics on operational efficiency. (CO1,K2) 10
- (4.2) Provide an in-depth analysis of the benefits of cloud computing in analytics. (CO1,K3) 10

5. Answer any one of the following:-

- (5.1) Describe inferential statistics and explain its key components.(CO2,K2) 10
- (5.2) Explain the importance of confidence intervals and their applications.(CO2,K2) 10

6. Answer any one of the following:-

- (6.1) Compare Z-score vs. IQR methods for outlier detection. (CO3, K4) 10
- (6.2) Compare Pandas vs. SQL for data manipulation.(CO3,K3) 10

7. Answer any one of the following:-

- (7.1) Discuss the advantages and disadvantages of supervised learning(CO4,K2) 10
- (7.2) Compare simple linear regression and multiple linear regression(CO4,K4) 10

8. Answer any one of the following:-

- (8.1) Describe types of logistic regression in detail.(CO5,K2) 10
- (8.2) Discuss application of Bayes theorem in classification.(CO5,K2) 10