

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

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

M.Tech

SEM: II - THEORY EXAMINATION (2025 - 2026)

Subject: Data Science for Security Analysis

Time: 3 Hours

Max. Marks: 70

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

15

1. Attempt all parts:-

- | | | |
|-------|---|---|
| (1.1) | E-commerce platforms enhance customer conversion rates by deploying: (CO1, K1) | 1 |
| | (a) manual spreadsheet inventories. | |
| | (b) static graphic flyers across web pages. | |
| | (c) recommender systems based on browsing habits. | |
| | (d) decentralized cryptographic hashes. | |
| (1.2) | Identify the terminology used to describe the difference between the observed value of the dependent variable and the predicted value. (CO1, K3) | 1 |
| | (a) Cost factor | |
| | (b) Bias | |
| | (c) Variance | |
| | (d) Residuals | |
| (1.3) | Select the specific visual element pair explicitly mentioned in the text that humans can quickly differentiate, such as distinguishing a red item from a blue item. (CO2, K2) | 1 |
| | (a) Text fonts | |
| | (b) Colors and patterns | |
| | (c) Grid spacing | |
| | (d) Column indices | |
| (1.4) | Primary use case for Graph NoSQL databases. (CO3, K1) | 1 |
| | (a) Generating static financial ledger reports | |
| | (b) Mapping highly interconnected networks like social media connections or fraud patterns | |

- (c) Storing simple sequential time-stamped log files
- (d) Backing up identical relational database tables
- (1.5) Structural arrangement defining an inverted file organization method. (CO4, K1) 1
 - (a) Documents arranged sequentially by their creation dates
 - (b) A list of relevant document records categorized under each indexing term
 - (c) A single flat file containing unindexed string data lines
 - (d) A strict hierarchical tree of connected table keys

2. Attempt all parts:-

- (2.1) Define the concept of 'Big Data' using its scale and structural boundaries. (CO1, K1) 2
- (2.2) Name the graphical curve-based method used to identify the optimum number of clusters (\$k\$) in K-means. (CO1, K1) 2
- (2.3) Define data blending. (CO2, K1) 2
- (2.4) Compare DDL and DML. (CO3, K2) 2
- (2.5) Define two rights of data owners. (CO4, K1) 2

SECTION-B

20

Answer any one of the following:-

- (3.1) Define 'Data Science' and list its five primary process steps in their exact sequential order. (CO1, K1) 4
- (3.2) Explain the operational profile of a 'Compromised Insider Threat' and differentiate its characteristics from a 'Malicious Insider Threat'. (CO1, K2) 4

Answer any one of the following:-

- (3.3) Explain the inverse relationship between bias error and variance error as described in the bias-variance trade-off framework, highlighting the consequences of failing to balance them. (CO1, K2) 4
- (3.4) Apply feature selection criteria and tree pruning strategies to a deep, highly complex decision tree model that exhibits near-zero training error but exceptionally high testing error. (CO1, K3) 4

Answer any one of the following:-

- (3.5) Identify the business communication medium designated in the text as the primary "go-to tool" for analyzing and sharing corporate data across organizational tiers, and list the target audience levels that must understand its impacts. (CO2, K3) 4
- (3.6) Define the exact terminology used by the author to describe the machine learning phenomenon where a model fits its first training dataset exceptionally well but fails to perform adequately on a fresh evaluation dataset. (CO2, K1) 4

Answer any one of the following:-

- (3.7) Explain relational database concept with structure of tables and relationships. (CO3, K2) 4
- (3.8) Explain graph database model using nodes and edges with suitable example. (CO3, K2) 4

Answer any one of the following:-

- | | | |
|--------|---|---|
| (3.9) | Explain concept of privacy in data science and describe its importance in protecting personal data. (CO4, K2) | 4 |
| (3.10) | Identify techniques used to reduce bias in algorithms and improve fairness in models. (CO4, K3) | 4 |

SECTION-C

35

4. Answer any one of the following:-

- | | | |
|-------|--|---|
| (4.1) | Explain the overall significance of Exploratory Data Analysis (EDA) in systemic modeling workflows, highlighting the necessity of data cleaning before algorithmic deployment. (CO1, K2) | 7 |
| (4.2) | Explain the fundamental concept of Data Loss Prevention (DLP) frameworks, explaining its structural enforcement across endpoint terminals and outbound network gateways. (CO1, K2) | 7 |

5. Answer any one of the following:-

- | | | |
|-------|--|---|
| (5.1) | Explain how data scientists evaluate the structural quality of clustering arrangements, focusing on the mathematical relationship between minimizing intracluster distance and maximizing intercluster distance. (CO1, K2) | 7 |
| (5.2) | Apply the formulas for Euclidean, Manhattan, and Minkowski distance metrics to compute a complete pairwise proximity matrix over a multi-dimensional coordinate array, proving how modifying the lambda parameter (λ) alters spatial separation. (CO1, K3) | 7 |

6. Answer any one of the following:-

- | | | |
|-------|--|---|
| (6.1) | Apply design principles of data visualization to construct an effective visualization framework for a dataset. (CO2, K3) | 7 |
| (6.2) | Explain data wrangling and analyze its importance in transforming raw data into usable datasets. (CO2, K2) | 7 |

7. Answer any one of the following:-

- | | | |
|-------|--|---|
| (7.1) | Explain limitations of relational databases in handling unstructured data. (CO3, K2) | 7 |
| (7.2) | Demonstrate advantages and challenges of NoSQL databases in large-scale systems. (CO3, K2) | 7 |

8. Answer any one of the following:-

- | | | |
|-------|---|---|
| (8.1) | Analyze regulatory frameworks such as GDPR and evaluate their contribution to ethical data practices. (CO4, K4) | 7 |
| (8.2) | Analyze causes and consequences of data breaches and assess strategies to minimize risks. (CO4, K4) | 7 |