# CBSE | DEPARTMENT OF SKILL EDUCATION

## DATA SCIENCE (SUBJECT CODE - 844)

### MARKING SCHEME FOR CLASS XII (SESSION 2024-2025)

**Max. Time: 2 Hours**                                            **Max. Marks: 50**

### General Instructions:

1. Please read the instructions carefully.
2. This Question Paper consists of **21 questions** in two sections – Section A & Section B.
3. Section A has Objective type questions whereas Section B contains Subjective type questions.
4. **Out of the given (5 + 16 =) 21 questions, a candidate has to answer (5 + 10 =) 15 questions in the allotted (maximum) time of 2 hours.**
5. All questions of a particular section must be attempted in the correct order.
6. **SECTION A - OBJECTIVE TYPE QUESTIONS (24 MARKS):**
    i. This section has 05 questions.
    ii. There is no negative marking.
    iii. Do as per the instructions given.
    iv. Marks allotted are mentioned against each question/part.
7. **SECTION B – SUBJECTIVE TYPE QUESTIONS (26 MARKS):**
    i. This section contains 16 questions.
    ii. A candidate has to do 10 questions.
    iii. Do as per the instructions given.
    iv. Marks allotted are mentioned against each question/part.

## SECTION A: OBJECTIVE TYPE QUESTIONS

| Q. No. | QUESTION | Source Material (NCERT/ PSSCIVE/ CBSE Study Material) | Unit/ Chap. No. | Page no. of source material | Marks |
|---|---|---|---|---|---|
| Q. 1 | **Answer any 4 out of the given 6 questions on Employability Skills (1 x 4 = 4 marks)** | | | | |
| i. | c. To whom? | NCERT | 1 | 17 | 1 |
| ii. | a. Intrinsic motivation | NCERT | 2 | 24 | 1 |
| iii. | a. Both Statement1 andStatement2 are correct | NCERT | 2 | 33 | 1 |
| iv. | b. A → 3; B → 1; C → 2 | NCERT | 3 | 55 | 1 |
| v. | b. Perseverance | NCERT | 4 | 103 | 1 |
| vi. | c. Dispose off the e-waste with the help of a certified partner | NCERT | 5 | 120 | 1 |
| Q. 2 | **Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)** | | | | |
| i. | a. Both A and R are correctand R is the correct explanation of A | NCERT | 1 | 2 | 1 |
| ii. | c. Outlier | NCERT | 2 | 11 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| iii. | a. Lee Goldman | NCERT | 3 | 16 | **1** |
| iv. | Cross Validation | NCERT | 4 | 25 | **1** |
| v. | d. the model is not a good fitand might need to be retrained. | NCERT | 5 | 35 | **1** |
| vi. | c. Recommendation systems | NCERT | 7 | 47 | **1** |
| Q. 3 | **Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)** | | | | |
| i. | c. Ethics | NCERT | 1 | 2 | **1** |
| ii. | c. Statement1 is correct but Statement2 is incorrect | NCERT | 2 | 8 | **1** |
| iii. | c. Supervised learningmodel | NCERT | 3 | 15 | **1** |
| iv. | b. Value of K is set foroptimum accuracy | NCERT | 4 | 23 | **1** |
| v. | a. The value of RSME is usually greater than or equal to the value of MAE | NCERT | 5 | 35 | **1** |
| vi. | a. Clustering | NCERT | 7 | 48 | **1** |
| Q. 4 | **Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)** | | | | |
| i. | d. HIPAA was passed in Indiato protect healthcare information from fraud and theft. | NCERT | 1 | 3 | **1** |
| ii. | Exploratory data analysis | NCERT | 2 | 7 | **1** |
| iii. | c. ii and iii | NCERT | 4 | 23 | **1** |
| iv. | d. Independent variable | NCERT | 5 | 34 | **1** |
| v. | b. Both Statement1 andStatement2 are incorrect | NCERT | 6 | 42 | **1** |
| vi. | c. Multi-class | NCERT | 7 | 48 | **1** |
| Q. 5 | **Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)** | | | | |
| i. | a. Both Statement1 andStatement2 are correct | NCERT | 1 | 2 | **1** |
| ii. | d. $1 \rightarrow b; 2 \rightarrow c ; 3 \rightarrow a$ | NCERT | 3 | 15 | **1** |
| iii. | Euclidean (or) Manhattan | NCERT | 4 | 25 | **1** |
| iv. | b. Continuous | NCERT | 5 | 37 | **1** |
| v. | d. Many, one | NCERT | 6 | 42 | **1** |
| vi. | b. Unsupervised learning needs human intervention. | NCERT | 7 | 47 | **1** |

# SECTION B: SUBJECTIVE TYPE QUESTIONS

| Q. No. | QUESTION | Source Material (NCERT/ PSSCIVE/ CBSE Study Material) | Unit/ Chap. No. | Page no. of source material | Marks |
|---|---|---|---|---|---|
| Answer any 3 out of the given 5 questions on Employability Skills in 20 – 30 words each (2 x 3 = 6 marks) | | | | | |
| Q. 6 | P in Respect stands for - Pay attention and focus on what thespeaker is saying. (1 mark for the word 'Pay'; and 1 mark for explanation) | NCERT | 1 | 5 | **2** |
| Q. 7 | Steps to overcome personality disorders • Talk to someone. Most often, ithelps to share your feelings. • Look after your physical health.A healthy body can help you maintain a healthy mind. • Build confidence in your abilityto handle difficult situations. • Engage in hobbies, such as music, dance and painting. Thesehave a therapeutic effect. • Stay positive by choosing wordslike 'challenges' instead of 'problems'. (Any 2 of the above – 1 mark each) | NCERT | 2 | 37 | **2** |
| Q. 8 | A worksheet is a collection of cells in the form of a grid (a network of lines that intersect each other, making rectangles). When you open a spreadsheet for the first time, you see a blank worksheet with the name 'Sheet1'. A workbook is a spreadsheet that has one or more worksheets. (1 mark each) | NCERT | 3 | 42 | **2** |
| Q. 9 | • Government Mudra Yojana • Credit Guarantee Scheme • Stand-Up India Scheme (Any 2 ; 1 mark each) | NCERT | 4 | 92 | **2** |
| Q. 10 | • increase the efficiency of energy and raw material. • reduce greenhouse gas emissions. • control waste and pollution. • protect and restore ecosystems. • support adaptation to the effects of climate change. ( Any 2 ; 1 mark each) | NCERT | 5 | 114 | **2** |

| | Answer any 4 out of the given 6 questions in 20 – 30 words each (2 x 4 = 8 marks) | | | | |
|---|---|---|---|---|---|
| **Q. 11** | Data privacy is the right of any individual to have control over how his or her personal information is collected and used.<br>(2 marks for relevant explanation) | NCERT | 1 | 3 | **2** |
| **Q. 12** | In the figure we can see that more than 2 variables are used.<br>Multivariate analysis is a more complex form of a statistical analysis technique and is used to analyze more than two variablesin the data set.<br>(1 mark for each point) | NCERT | 2 | 10 | **2** |
| **Q. 13** | • K-NN does not have a specific training phase where it learns about the data.<br>• It usesall the training data while performing a classification operation.<br>Hence follows the principle of lazy learning.<br>(1 mark for each point) | NCERT | 4 | 24 | **2** |
| **Q. 14** | The given picture is **Decision Tree**<br>Benefits of Decision Tree<br>1. Decision Trees are versatile as they can be used toany kind of problem at hand - classification or regression.<br>2. Decision tree outputs arevery easy to understand even forpeople from a non-analytical background.<br>3. It can handle both numerical and categoricalvariables.<br>4. Decision trees are a good tool for classifying observations when the trend is non-linear<br>(1 mark for identifying decision tree; any 2 benefits – ½ mark each) | NCERT | 3 | 16 | **2** |
| **Q. 15** | We usually use linear regression when we want to know:<br>• nature and strength relationship between two variables (e.g., is there a positive relationship between rainfall and crop growth).<br>• The predicted value of the dependent variable for a given value of the independent variable (e.g. the amount of crop growth for a certain level of rainfall).<br>(1 mark per each point) | NCERT | 5 | 34 | **2** |
| **Q. 16** | Unsupervised algorithms discover hidden patterns or data groupings with their ability to discover similarities and differences in the given unlabeled dataset.<br>(2 marks for any relevant explanation) | NCERT | 7 | 47 | **2** |

**Answer any 3 out of the given 5 questions in 50– 80 words each (4 x 3 = 12 marks)**

| Q. 17 | Data governance can be thought of as a collection of people, technologies, processes, and policies that protect and help to manage efficient use of data. The following points are to be kept in mind while dealing about data.<br>• Keep the data secure.<br>• Create machine learning models that are impartial androbust<br>• Be as open and accountableas possible<br>• Use technologies and data architecture that has the minimum intrusion necessary.<br>(1 mark for defining the term data governance; any 3 benefits – 3 marks) | NCERT | 1 | 2 | **4** |
|---|---|---|---|---|---|
| Q. 18 | Steps of data cleaning are as mentioned below.<br>• Remove duplicate observations -<br>• Remove irrelevant observations<br>• Remove unwanted outliers<br>• Fix data type issues<br>• Handle missing data<br>(any 4 points; 1 mark per point) | NCERT | 2 | 10, 11 | **4** |
| Q. 19 | Steps to create a decision tree are:<br>1. The main decision that you are trying to make should be placed at the very top of the decision tree. Therefore, the main objective should be the "root" of the entire diagram.<br>2. Next, you need to draw the branches and leaf nodes. For every possible decision, stemming from the root make a branch. One root or node can have two or more branches. At the end of the branches, attach the leaf nodes. The leaf nodes should represent the results of each decision. If another decision has to be made, draw a square leaf node. If the outcome is not quite certain, you should draw a circular node<br>3. Finally, you need to calculate the probability of success of each decision being made. While creating the decision tree, it is essential to do some research, so that you can predict the probability of each decision. To do this research, you may examine old data or assess previous projects. Once you calculate the expected value of each decision in tree, put the values on the branches<br>(1 mark per point; 1 mark for any suitable example) | NCERT | 3 | 17 | **4** |

| Q. 20 | a. Multiple Linear Regression<br>b. Multiple Linear Regression uses multiple independent variables to predict the outcome of a dependent variable. For example, effects of age, weight, and height on cholesterol levels of an individual. Here, age, weight, and height are independent variables, and cholesterol level is the dependent variable because it is dependent on the factors age, height, and weight.<br>c. The basic model of multiple linear regression is $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \epsilon_i$ where i = 1, 2, …, n foreach observation.<br>(**a:** 1 mark<br>**b:** 2 marks<br>**c:** 1 mark) | NCERT | 6 | 42 | **4** |
|---|---|---|---|---|---|
| Q. 21 | The k-means clustering method is a technique that is used to spot clusters of data classes in a dataset.<br> **(½ mark for definition)**<br>The k-means clustering algorithm works is as follows:<br>1. First, specify the number of clusters K depending on the input**. (½ mark)**<br>2. Initialize the centroids by shuffling the data points and then selecting K data points forthe centroids randomly.<br>**(½ mark)**<br>3. Do iterations till the point that there is no change to the centroids such that the assignment of data points to clusters isn't changing.<br>**(½ mark)**<br>4. Compute the sum of the squared distance between data points and all centroids.<br>**(1 mark)**<br>5. Classify or mark each data point to the cluster it is closest to.<br>**(1 mark)** | NCERT | 7 | 49 | **4** |