

<b>Roll No.</b>							
-----------------	--	--	--	--	--	--	--

21554

**BCA V SEMESTER [MAIN/A.T.K.T.] EXAMINATION**  
**FEBRUARY - 2022**

# INTRODUCTION TO DATA SCIENCE

**[Max. Marks : 85]**

*[Time : 3:00 Hrs.]*

**[Min. Marks : 28]**

**Note : All THREE Sections are compulsory. Student should not write any thing on question paper.**  
नोट : सभी तीन खण्ड अनिवार्य हैं। विद्यार्थी प्रश्न-पत्र पर कुछ न लिखें।

**[Section - A]**

This Section contains **Multiple Choice Questions**. Each question carries **1 Mark**.

इस खण्ड में बहविकल्पीय प्रश्न हैं। प्रत्येक प्रश्न 1 अंक का है।

- Q. 01** Which of the following are the Data source in Data Science ?  
 a) Structured                                      b) Unstructured  
 c) Both (a) and (b)                            d) None of these  
 डेटा साइन्स में डेटा स्रोत निम्न में से कौन से हैं –  
 a) संरचित    b) असंरचित  
 c) (a) और (b) दोनों                         d) उपरोक्त में से कोई नहीं
- Q. 02** In 1991, R was developed in the department of statistics at the university of -  
 a) John Hopkins                                 b) California  
 c) Harvard                                         d) Auckland  
 1991 में R को \_\_\_\_\_ विश्वविद्यालय में सांख्यिकी विभाग में बनाया गया था -  
 a) जॉन हॉपकिंस                                b) कैलिफोर्निया  
 c) हार्वर्ड                                         d) ऑकलैंड
- Q. 03** \_\_\_\_\_ may be defined as the data objects that do not comply with the general behavior or model of the data available -  
 a) Outlier analysis                                b) Evolution analysis  
 c) Prediction                                      d) Classification  
 \_\_\_\_\_ को डेटा ऑब्जेक्ट के रूप में परिभाषित किया जा सकता है जो उपलब्ध डेटा के सामान्य व्यवहार या मॉडल का अनुपालन नहीं करता है -  
 a) बाहरी विश्लेषण                              b) विकास विश्लेषण  
 c) भविष्यवाणी                                  d) वर्गीकरण
- Q. 04** How many types are available in machine learning ?  
 मशीन लर्निंग में कितने प्रकार उपलब्ध हैं –  
 a) 1    b) 2  
 c) 3    d) 4

P.T.O.

**Q. 05** How do you choose right node while constructing a decision tree -

- a) An attribute having high entropy      b) An attribute having high entropy and information gain  
c) An attribute having the lowest information gain      d) An attribute having the highest information gain.  
निर्णय वृक्ष का निर्माण करते समय आप सही नोड कैसे चुनते हैं –  
a) उच्च एन्ट्रॉपी वाली विशेषता      b) उच्च एन्ट्रॉपी और सूचना लाभ वाली विशेषता  
c) सबसे कम सूचना लाभ वाली विशेषता      d) उच्चतम सूचना लाभ वाली विशेषता

---

**[Section - B]**

This section contains **Short Answer Type Questions**. Each question carries **5 Marks**.

इस खण्ड में लघुउत्तरीय प्रश्न हैं। प्रत्येक प्रश्न 5 अंकों का है।

---

**Q. 1** Define Data Science ? Write down applications of Data Science.

डेटा विज्ञान को परिभाषित कीजिये। डेटा विज्ञान के अनुप्रयोगों को लिखिये।

**OR**

Define Histogram ? Explain key difference between histogram and bar graph.

हिस्टोग्राम को परिभाषित कीजिये। हिस्टोग्राम तथा बार ग्राफ के बीच अंतर बतलाइये।

**Q. 2** Explain data types in R programming with example.

R प्रोग्रामिंग में डेटा टाईप्स को उदाहरण सहित समझाइये।

**OR**

Write down the key difference between Matrix and Data Frame in R.

R में मैट्रिक्स तथा डेटा फ्रेम के मध्य अंतर बतलाइये।

**Q. 3** Define Outliers ? Explain most common causes of outliers on a data set.

आउटलेर्स को परिभाषित कीजिये ? डेटा सेट में आउटलेर्स के सामान्य कारणों को समझाइये।

**OR**

Write down the benefit of data cleansing ?

डेटा क्लीनिंग के लाभ लिखिये।

Cont. . . .

- Q. 4** What is Machine Learning ? How does machine learning work ?  
मशीन लर्निंग क्या है ? मशीन लर्निंग कैसे काम करती है ?

**OR**

Explain Ridge Regression ?  
रिज रिग्रेशन की व्याख्या कीजिये ?

- Q. 5** What do you mean by Boosting ?  
बूस्टिंग से आप क्या समझते हैं ?

**OR**

Explain Logistic Regression ?  
लॉजिस्टिक रिग्रेशन को समझाइये ?

---

**[Section - C]**

This section contains **Essay Type Questions**. Each question carries **11 marks**.  
इस खण्ड में दीर्घउत्तरीय प्रश्न हैं। प्रत्येक प्रश्न 11 अंकों का है।

---

- Q. 6** Explain the data science process in detail.  
डेटा साइन्स प्रोसेस को विस्तार से समझाइये ?

**OR**

What is Quantitative and Qualitative data ? Explain difference between quantitative and qualitative data ?  
मात्रात्मक एवं गुणात्मक डेटा क्या है ? मात्रात्मक और गुणात्मक डेटा के बीच अंतर स्पष्ट कीजिये ?

- Q. 7** Explain following data structure in R with example -  
R में निम्नलिखित डेटा स्ट्रक्चर को उदाहरण सहित समझाइये -  
i) Vector  
ii) Data Frame.

**OR**

Explain loops in R with example.  
R में लूप्स को उदाहरण सहित समझाइये।

- Q. 8** What is Feature Engineering ? Explain process and need of feature engineering.  
फीचर इंजीनियरिंग क्या है ? फीचर इंजीनियरिंग की प्रक्रिया तथा आवश्यकता की व्याख्या कीजिये।

**OR**

**P.T.O.**

Define Data Cleansing ? Explain the process of data cleansing in detail.

डेटा क्लीजिंग को परिभाषित कीजिये। डेटा क्लीजिंग की प्रक्रिया को विस्तार से समझाइये।

**Q. 9** Explain the life cycle of machine learning ?

मशीन लर्निंग के जीवन चक्र को समझाइये ?

**OR**

What is Regression ? Explain simple and multiple linear regression.

रिग्रेशन क्या है ? सरल तथा बहु रेखीय रिग्रेशन को समझाइये।

**Q. 10** What is Over fitting ? Explain how to detect and prevent over fitting.

ओवर फिटिंग क्या है ? ओवर फिटिंग का पता लगाने एवं उसे रोकने का तरीका समझाइये।

**OR**

Explain Decision Tree algorithm with example.

उदाहरण के साथ डिसीज़न ट्री एल्गोरिथम को बतलाइये।

\_\_\_\_\_○\_\_\_\_\_