

Total No. of Printed Pages: 2

B.Sc Course (CBCS) Ordinance Sem-VI  
EXAMINATION APRIL 2024  
COMPUTER SCIENCE - INTRODUCTION TO DATA ANALYTICS

[Time: 2:00 Hours]

[Max. Marks: 60]

**Instructions:** 1. All questions are compulsory  
2. Figures to the right indicate marks

**Q1** Answer any five from the following.

- a. Define the term Big Data.
- b. Differentiate between population and sample.
- c. What is the purpose of conducting exploratory data analysis?
- d. Why is there a need to perform feature selection?
- e. What is clustering? List any two applications of it.
- f. What is a community in a social network graph?
- g. What is a heat map?

2x5=10

**Q2** Answer any five from the following.

- a. List any two real time applications of Data Science.
- b. Define the term Probability. Give a suitable example.
- c. Differentiate between univariate and bivariate data. Give example of each.
- d. What do you understand by the term User Retention?
- e. List any two machine learning algorithms and its applications.
- f. How is data visualization used in machine learning?
- g. Explain the term Data Ownership with respect to Data Science.

2x5=10

**Q3** Answer A or A. B is compulsory

A. Explain market segmentation with the help of a suitable example.

5

**OR**

A. With the help of a suitable example explain the working of Poisson distribution.

5

B. Explain the Data Science process.

5

**Q4** Answer A or A. B is compulsory

A. What is filter selection method? Explain its advantages and disadvantages.

5

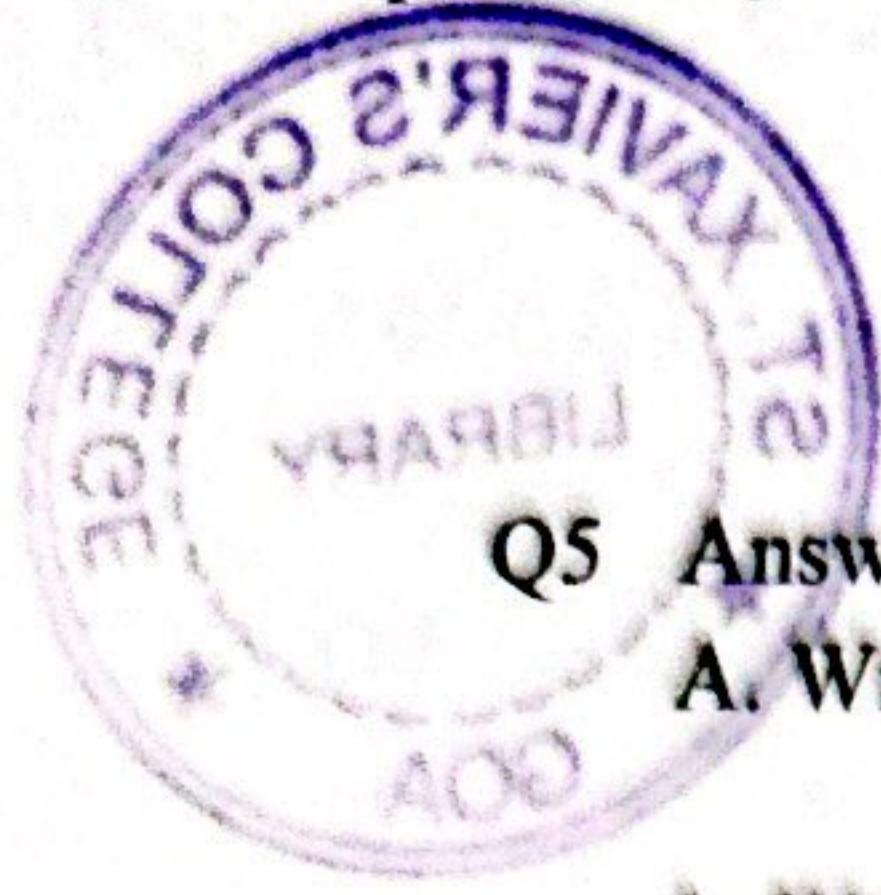
**OR**

A. How can feature selection optimize the modeling process?

5

B. What is exploratory data analysis? And why do we conduct it before the process of modeling?

5



**Q5 Answer A or A. B is compulsory**

A. Write a note on K-means machine learning algorithm.

**OR**

A. Write a note on decision tree machine learning algorithm.

B. Differentiate between linear regression and logistic regression.

**Q6 Answer A or A. B is compulsory**

A. List and explain the neighbourhood properties of graphs.

**OR**

A. Explain any five principles used in data visualization.

B. What is social cooling? Explain how it works?

(6) Total No. of Printed Pages: 2

T.Y.B.Sc. Course (CBCS) Ordinance (Sem. V)

EXAMINATION May 2023

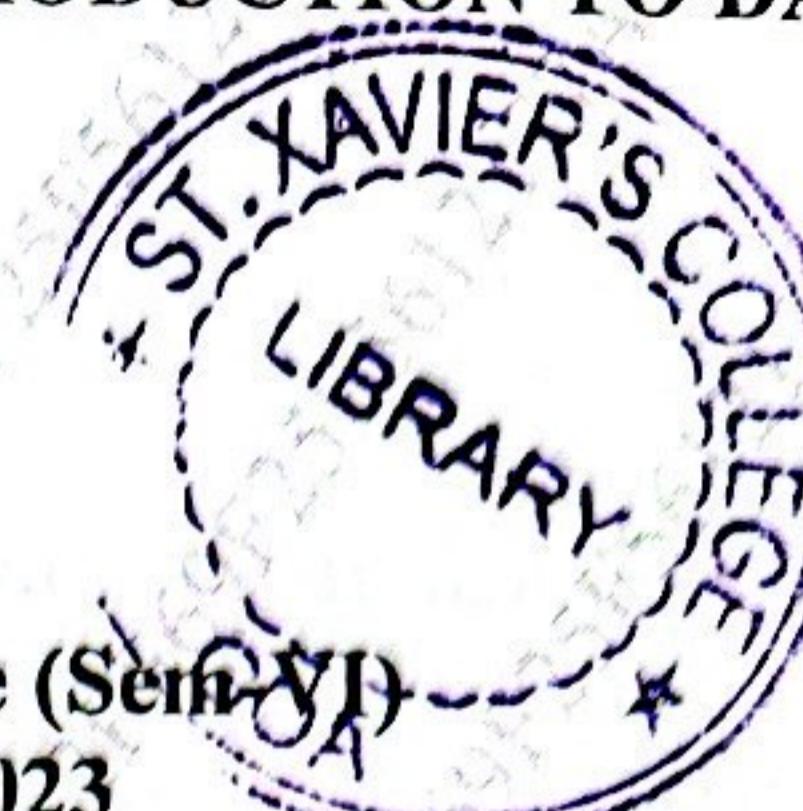
COMPUTER SCIENCE - INTRODUCTION TO DATA ANALYTICS

(6)

[Time:2 Hours]

[Max. Marks:60]

Instructions: 1. All questions are compulsory  
2. Figures to the right indicate marks



(6) Q1 Answer any five from the following.

(2x5=10)

- What is data science?
- Define statistical inference.
- What is exploratory data analysis?
- What is dimensionality reduction?
- Differentiate between simple linear regression and multiple linear regression.
- How can directed graph be used to model a network?
- What is data visualization in machine learning?

(6) Q2 Answer any five from the following.

(2x5=10)

- What is datafication?
- Differentiate between probability mass function and probability density function.
- What is the purpose of conducting exploratory data analysis?
- Why is there a need to select a subset of features from the entire data set?
- State the limitations of using decision tree for classification problems.
- What is a heatmap?
- State the similarities between bernoulli and binomial distribution.

(6) Q3 A. Explain survival analysis with the help of a suitable example.

(5)

OR

- Differentiate between population and sample. Explain with a suitable example.
- Explain the current landscape of perspective of data science.

(5)

(5)

**Q4** A. Explain feature selection using filter approach.

**OR**

A. State the advantages and disadvantages of wrapper approach.

B. With the help of a suitable example explain box plot used during exploratory data analysis.

**Q5** A. Write a note on linear regression.

**OR**

A. Write a note on random forest.

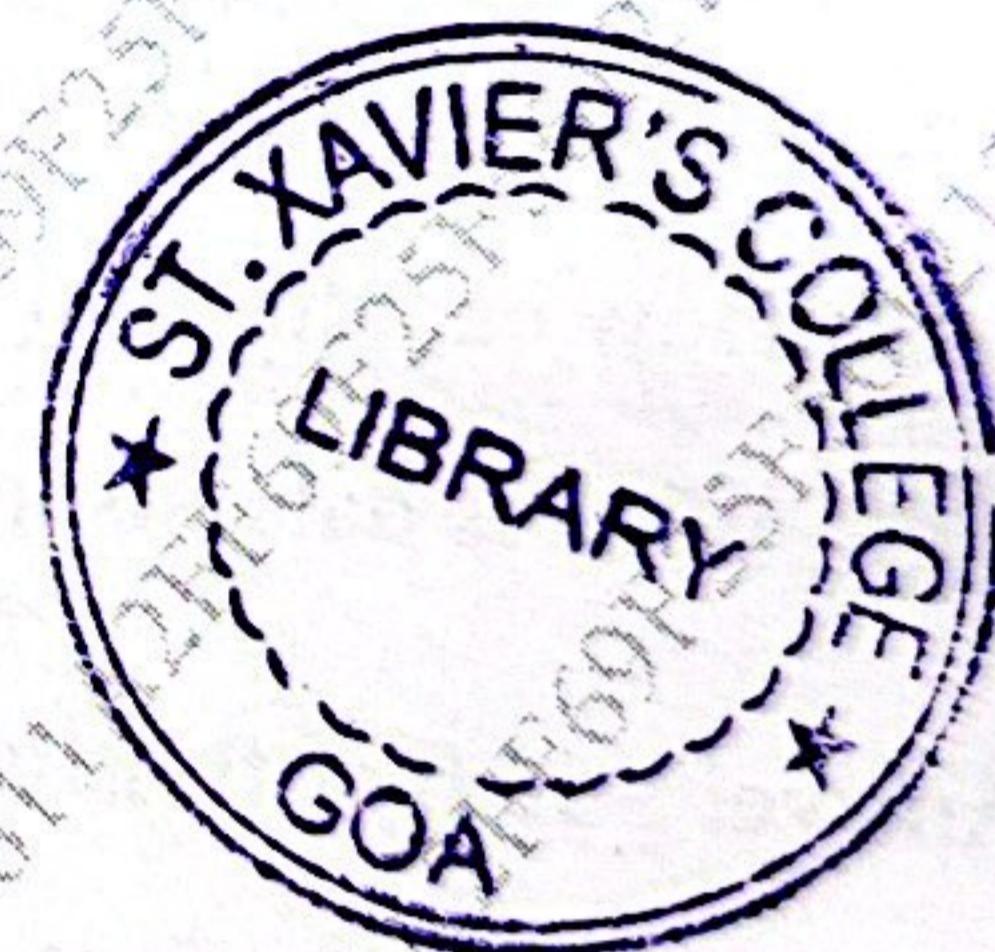
B. What is a social network? State the characteristics of social network.

**Q6** A. What is the importance of data visualization in machine learning?

**OR**

A. With the help of a suitable example, explain uniform distribution.

B. Write a note on ethical issues in data science.



**ST. XAVIER'S COLLEGE, MAPUSA GOA  
SIXTH SEMESTER END EXAMINATION  
JUNE 2022 (REGULAR/REPEAT)**



**T.Y.B.SC OC – 66 CBCS**

**SUBJECT: COMPUTER SCIENCE**

**Subject Code: UCSD107**

**Paper Title: Introduction to Data Analytics**

**DURATION : Two Hours**

**MAX. MARKS : 60**

---

**Instructions :**

- 1) Questions 1 – 6 are compulsory. However, internal choice is available.
- 2) Figures to the right indicate marks.

**Q.1) Answer any five of the following:**

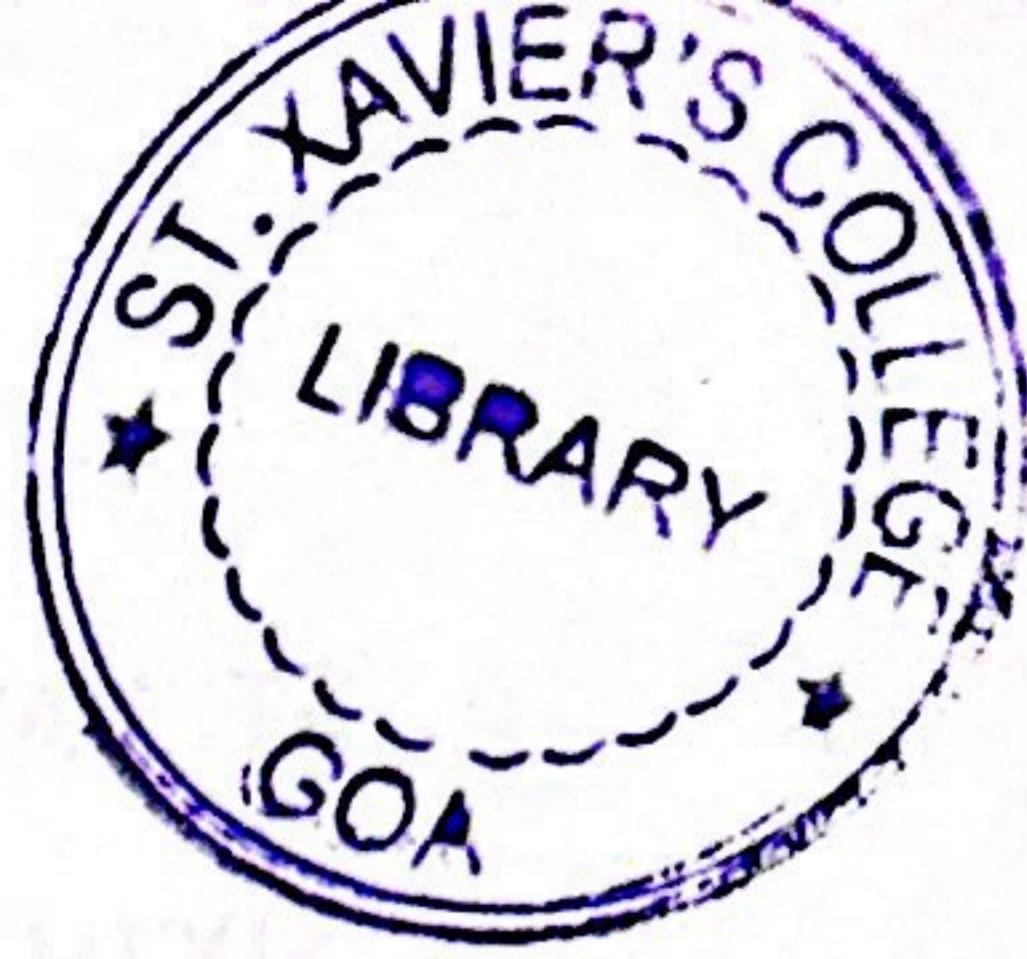
**(5x2=10)**

- a) What do you understand by the term datafication?
- b) Distinguish between population and sample.
- c) What is the purpose for conducting exploratory data analysis (EDA)?
- d) Why is there a need to perform feature selection?
- e) Differentiate between supervised and unsupervised machine learning algorithms.
- f) Why is data visualization important in data science?
- g) Differentiate between classification and clustering machine learning algorithms.

**Q.2) Answer any five of the following:**

**(5x2=10)**

- a) Differentiate between univariate, bivariate and multivariate data and its analysis? Give an example of each.
- b) State any two advantages and disadvantages of Wrapper selection method.
- c) What is classification? Give any two applications of it.
- d) State any two uses of Logistic regression.
- e) Give any two applications of Data Science.
- f) List any two examples where Statistical Modeling is used.
- g) Differentiate between discrete data and categorical data.



**Q.3 Answer the following:**

A) Explain the working of Poisson distribution.

OR

A) Explain the working of Exponential distribution.

05

B) Explain the Data Science Process.

05

05

**Q.4 Answer the following:**

A) List and explain any two types of Plots used for EDA.

05

OR

A) List and explain any two Statistical tools used for EDA.

05

B) What is a Filter selection method. State its advantages and disadvantages.

05

**Q.5 Answer the following:**

A) With the help of a suitable example explain Linear Regression. State any two applications of it.

05

OR

A) With the help of a suitable example explain Decision Trees. State any two applications of it.

05

B) Write a note on K-means algorithm.

05

05

**Q.6 Answer the following:**

A) What is Social Cooling? Explain how it works?

05

OR

A) List and explain the negative effects of Social Cooling.

05

B) List and explain any five principles of Data Visualization.

05

05

ALL THE BEST