

Roll No



**PRESIDENCY UNIVERSITY
BENGALURU**

SET A

**SCHOOL OF MANAGEMENT
END TERM EXAMINATION - JAN 2024**

Semester : Semester V - 2021
Course Code : BBB3004
Course Name : Machine Learning
Program : BBA

Date : 09-JAN-2024
Time : 1:00 PM - 4:00 PM
Max Marks : 100
Weightage : 50%

Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and non-programmable calculator are permitted.
- (iv) Do not write any information on the question paper other than Roll Number.

PART A

ANSWER ALL THE QUESTIONS

5 X 2M = 10M

1. How is data important for Machine Learning?
(CO1) [Knowledge]
2. List some strengths of Python as a programming language.
(CO2) [Knowledge]
3. What are the main techniques that improve data quality before analysis?
(CO3) [Knowledge]
4. Interpret correlation matrix to understand the connections between variables?
(CO4) [Knowledge]
5. Explain the basic function of an algorithm, and how does it contribute to solving problems in computer science?
(CO5) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS

5 X 10M = 50M

6. Explain the key libraries and frameworks that make Python a preferred choice for data science?
(CO1) [Comprehension]

7. Write a Python program to calculate grades based on a student's score. The grading conditions are as follows:

- If the score is 90 or higher, the grade is "A."
- If the score is between 80 and 89 (inclusive), the grade is "B."
- If the score is between 70 and 79 (inclusive), the grade is "C."
- If the score is between 60 and 69 (inclusive), the grade is "D."
- If the score is below 60, the grade is "F."

Use IF, ELIF, ELSE Statements.

(CO2) [Comprehension]

8. Describe the key phases in data analysis, and how do they contribute to understanding and interpreting data effectively?

(CO3) [Comprehension]

9. Demonstrate five common types of data visualizations that can be created using Python.

(CO4) [Comprehension]

10. Describe the Random Forest algorithm and its role in machine learning. Elaborate on the fundamental principles guiding its ensemble learning approach.

(CO5) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS

2 X 20M = 40M

11. How does the choice of standardization technique impact the performance of machine learning algorithms, especially when dealing with datasets with varying scales and distributions?

(CO3) [Application]

12. You have been provided with a dataset containing examples of emails, each labeled as either spam or ham (non-spam). The dataset consists of the email content and the corresponding category. Your task is to build a Naive Bayes classifier to predict whether a given email is spam or ham based on its content.

Email Text	Category
Claim your exclusive offer today!	Spam
Team meeting agenda for tomorrow attached.	Ham
You've won a lottery! Contact us to claim your prize.	Spam
Reminder: Project presentation at 3 PM.	Ham
Limited-time discount for our loyal customers.	Spam

Questions:

1. How would you preprocess the email text for building the Naive Bayes classifier? List the steps involved.
2. Calculate the prior probabilities $P(\text{Spam})$ and $P(\text{Ham})$ using the Provided Dataset
3. Compute the conditional probabilities $(\text{word}|\text{Spam})$ $P(\text{word}|\text{Spam})$ and $(\text{word}|\text{Ham})$

(CO5) [Application]