

Roll No																			
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SCHOOL OF ENGINEERING  
MID TERM EXAMINATION - OCT 2023**

**Semester :** Semester VII - 2020

**Course Code :** CSE3133

**Course Name :** Sem VII - CSE3133 - Predictive Analytics for Big Data

**Program :** COD

**Date :** 30-OCT-2023

**Time :** 11:30AM - 1:00PM

**Max Marks :** 60

**Weightage :** 30%

---

**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 2 = 10M)**

1. Find out the characteristics of big data.

(CO1) [Knowledge]

2. Define the data volume.

(CO1) [Knowledge]

3. Write down the difference between classical and exploratory data analysis.

(CO2) [Knowledge]

4. Define stratified sampling.

(CO2) [Knowledge]

5. Show the architecture diagram for EDA.

(CO2) [Knowledge]

**PART B**

**ANSWER ALL THE QUESTIONS**

**(2 X 15 = 30M)**

6. Summarise the big data technologies in
- a. Apache Hadoop
  - b. Apache Spark
  - c. Mongo DB
  - d. Cassandra
  - e. Apache Kafka
  - f. Qlikview
  - g. Qlik Sense
  - h. Tableau

(CO1) [Comprehension]

7. Explain the Big Data Infrastructure with its design architecture diagram, where it accepts data in various formats, allowing real-time storage and analysis.

(CO1) [Comprehension]

**PART C**

**ANSWER THE FOLLOWING QUESTION**

**(1 X 20 = 20M)**

8. Solve the weight of evidence and information value for the given age-group table and finally make variable predictiveness

Age-Group	Good	Bad	WoE	Information
18-25	1000	170		
25-35	1000	230		
36-45	1800	95		
46-55	1602	106		
56-70	1900	92		

(CO2) [Application]