

Total No. of Questions : 5]

SEAT No. :

PB-1464

[Total No. of Pages : 2

[6226]-306

S.Y. B.B.A. (C.A.)

CA - 305 : BIG DATA

(2019 Pattern) (CBCS) (Semester - III)

Time : 2½ Hours]

[Max. Marks : 70

Instructions to the candidates:

- 1) *All questions are compulsory.*
- 2) *Figures to the right indicate full marks.*

Q1) Attempt any EIGHT of the following.

[8 × 2 = 16]

- a) What is big data?
- b) What is supervised learning?
- c) What is population?
- d) Define sample.
- e) What is WEKA?
- f) What is KNN?
- g) Define SVM.
- h) What is the use of histogram?
- i) What is EM algorithm?
- j) What is R?

Q2) Attempt any FOUR of the following.

[4 × 4 = 16]

- a) Explain the types of Data Analytics.
- b) What is probability? Explain its types.
- c) Explain Association rule mining.
- d) Explain applications of Apriori algorithm.
- e) Explain types of cluster analysis.

P.T.O

Q3) Attempt any FOUR of the following.

[4 × 4 = 16]

- a) Explain types of regression models.
- b) What is digital data? Explain its types.
- c) Explain five applications of Machine Learning.
- d) State advantages and disadvantages of SVM.
- e) Explain Decision tree with example.

Q4) Attempt any FOUR of the following.

[4 × 4 = 16]

- a) Explain Naive Bayes with the help of example.
- b) Explain applications of big data.
- c) Write an R program to sort vector in ascending and descending order.
- d) Write an R program to create a simple bar plot of five subject's marks.
- e) Write an R program accept any year is input and check whether the year is leap or not.

Q5) Write a short note on any TWO of the following.

[2 × 3 = 6]

- a) Data Science.
- b) Data Visualization.
- c) Data Frames.

