Total No. of Questions : 5]

SEAT No. :

**PD-1440**

[Total No. of Pages : 2

# [6441] - 506
# S.Y. B.B.A. (CA)
# 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 right indicate marks.***

**Q1)** Attempt any Eight of the following : **[8 × 2 = 16]**

a) What are the five's of big data?

b) What is unsupervised learning?

c) What is regression?

d) What is sample?

e) What is pipe operator?

f) What is the statistical inference?

g) Enlist stages of Data Science.

h) What is digital data?

i) What is probability?

j) What is R?

***P.T.O.***

**Q2)** Attempt any FOUR of the following : **[4 × 4 = 16]**

a) Explain the life cycle of data analytics.

b) Explain probability distribution modeling.

c) Describe Association rule for mining.

d) Explain Apriori algorithm.

e) Explain K-means of clustering algorithm.

**Q3)** Attempt any FOUR of the following : **[4 × 4 = 16]**

a) Explain correlation with its type.

b) Explain tools used in big data.

c) Explain the steps in Machine Learning.

d) Explain SVM algorithm in detail.

e) Differentiate between unstructured digital data and semi structured digital data.

**Q4)** Attempt any FOUR of the following : **[4 × 4 = 16]**

a) How Naive Bayes algorithm works.

b) Explain applications of big data in E- Commerce

c) Write an R program to calculate binary into decimal of a given vector.

d) Write an R program to find sum, mean and product of a vector.

e) Write an R program to find maximum and minimum value of a given vector.

**Q5)** Write a short note on Any TWO of the following : **[2 × 3 = 6]**

a) Data Manipulation.

b) Packages in R.

c) Loops in R