Total No. of Questions : 5]

SEAT No. :

PA-1034

[Total No. of Pages : 2

[5902]-64

T.Y. B.Sc. (Semester - VI)

COMPUTER SCIENCE

CS-364 : Data Analytics

(2019 Pattern) (CBCS)

Time : 2 Hours] *[Max. Marks : 35*

Instructions to the candidates:

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

Q1) Attempt any eight of the following (out of 10). **[8 × 1 = 8]**

a) Define Data Analytics.

b) What is AVC & ROC curve?

c) Write any two applications of Supervised Machine Learning.

d) Give the formula for support & confidence.

e) What is an outlier?

f) State applications of NLP.

g) What is web scraping?

h) What is the purpose of n-gram?

i) Define classification.

j) Define Recall.

Q2) Attempt any four of the following (Out of five). **[4 × 2 = 8]**

a) Explain the concept of underfitting & overfitting.

b) What is linear Regression? What type of Machine learning applications can be solved with linear Regression?

P.T.O.

c) What is Social Media Analytics?

d) What are the advantages of FP-growth Algorithm?

e) What are dependent & independent variables?

Q3) Attempt any two of the following (Out of three). **[2 × 4 = 8]**

a) What are frequent itemsets & association rules? Describe with example.

b) What is stemming & lemmatization?

c) Explain various types of Data Analytics.

Q4) Attempt any two of the following (Out of three). **[2 × 4 = 8]**

a) What is Bag of words & DOS tagging in NLP?

b) What is Logistic Regression? Explain it with example.

c) Consider the following database & find out the frequent itemset using Apriori Algorithm with minimum support threshold = 3.

T. id.	Item purchased
1	M,T,B
2	E,T,C
3	M,E,T,C
4	E,C
5	J

Q5) Attempt any one of the following (Out of 2). **[1 × 3 = 3]**

a) Define the terms

i) Confusion Matrix

ii) Accuracy

iii) Precision

b) What is Machine Learning? Explain its type.