

B.E/B.TECH. Degree Examination, December 2020

Semester - VI

IT16002 - Data Science Using Python

(Regulation 2016)

Time: Three hours

Maximum : 80 Marks

Answer **ALL** questions**PART A - (8 X 2 = 16 marks)**

1. A tuple $t = (10, 20, 45, 36)$.Identify the incorrect statement below:
 - a) `print(t[-4:0])`
 - b) `t[3] = 55`
 - c) `print(max(t))`
 - d) `print(len(t))`
2. Two random variables, A and B, where $P(A) = 1/3$, $P(B) = 1/2$, and $P(A | \neg B) = 1/4$, what is $P(A | B)$?
 - a) $1/6$
 - b) $1/4$
 - c) $3/4$
 - d) 1
3. x is a sample of single value. For the following equations,

1) $y = ax + e$

2) $y = ax + bx^2 + e$

Which one of the regression models is more appropriate to fit the training data?

- a) Linear model
 - b) Multiple Regression model
 - c) both will equally fit
 - d) not enough data
4. Which of the following is not an impurity measure?
 - a) Gini Index
 - b) Information Gain
 - c) Entropy
 - d) Pruning

5. What will be the output of the following Python code?

```
A = [[1, 2, 3],  
     [4, 5, 6],  
     [7, 8, 9]]  
A[1]
```

6. Is correlation influenced by outliers? Justify

7. Differentiate overfitting and underfitting of data.

8. Compare content based and collaborative filtering.

PART B - (4 X16 = 64 marks)

09. (a) Write a python program to simulate conditional probability and bayes theorem. (16)

(OR)

(b) (i) Differentiate correlation and causation. (8)

(ii) Illustrate with relevant python code the mean, median and mode. (8)

10. (a) Analyze the process of getting credentials and accessing services using python. (16)

(OR)

(b) Illustrate with relevant python code the concept of Dimensionality reduction. (16)

11. (a) Write a python code to implement K nearest neighbor classifier with a suitable dataset. (16)

(OR)

(b) Simulate a simple linear regression model using a python code. (16)

12. (a) For the 1- dimensional data set {57,15,45,60}, perform hierarchical clustering – single linkage and complete linkage. Illustrate the clustering algorithm using a python code. (16)

(OR)

(b) Analyze the training of feed forward neural network with relevant python codes. (16)