CVR COLLEGE OF ENGINEERING

An UGC Autonomous Institution - Affiliated to JNTUH

B.Tech. **IV** Year **I** Sem. **Substitution Examinations**, December - 2021

Subject**: Fundamentals of Data Science (OE-1)**

Date: 23-12-2021(FN) Time: 1 hr 45 min Max. Marks: **40 M**

PART – A

Answer **ALL** questions **5x2 = 10 M**

1. Compare discrete and continuous data. L1 (CO1)
2. How do you construct a histogram? L3 (CO2)
3. Give the relationship between Y and X in linear regression. L4 (CO3)
4. Explain recommendation system. L2 (CO4)
5. How decision tree is constructed? L3 (CO5)

PART – B

Answer **ALL** questions **3x10 = 30 M**

6. (a) What is Data Science and why Data Science is in demand? L1 (CO1)
   (b) A factory has two machines A and B. Past records show that machine Aproduces 30% of the total output and machine B the remaining 70%. Machine A produces 5% defective articles and machine B produce 1% defective items. An item is drawn at random from the output and found to be defective. What is the probability that it was produced by machine-B? (5+5) L4 (CO1)

(OR)

7. (a) The following information pertaining the wags of 30 workers in a factory (in hundreds)33, 32, 55, 47, 21, 50, 27, 12, 68, 49, 40, 17, 44, 48, 62,24, 33, 42, 38, 45, 26, 33, 44, 48, 52, 30, 58, 37, 38, 35. Draw the Stem-and-Leaf-Plot. (5+5) L4 (CO2)
   (b) For the data in 7(a), find quartiles and draw Box-Whisker Plot. L4 (CO2)

8. (a) Explain linear regression algorithm in machine learning. (4+6) L2 (CO3)
   (b)Playing cricket(Y) depends on outlook(X_1), temperature(X_2), humidity(X_3) and wind(X_4). The data on these variables are given in the following table: L4 (CO3)

Day	X_1	X_2	X_3	X_4	Y
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes
5	Rain	Cool	Normal	Strong	No
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Sunny	Cool	Normal	Weak	Yes
9	Rain	Mild	Normal	Weak	Yes
10	Overcast	Mild	High	Strong	Yes
11	Overcast	Hot	Normal	Weak	Yes
12	Rain	Mild	High	Strong	No

Find Y when X_1=Sunny, X_2=Cool, X_3= High, X_4= Strong using Naive Bayes method.

P.T.O

(OR)

9. (a)Give the singular value decomposition algorithm (4+6) L1 (CO4)
   (b)Perform the principal component analysis to the following data: L4 (CO4)

X_1	3	10	4	8	5
X_2	10	8	9	8	5

10. (a) Write about the role of Domain experts in Data Science? (5+5) L3 (CO5)
    (b) Write about Feature Selection methods In Data Science? L3 (CO5)

(OR)

11. (a) Briefly explain random forests (4+6) L2 (CO5)
    (b) Draw the decision tree to the data given in Question 8(b). L2 (CO5)

*****