(Pages : 3) **N – 6759**

Reg. No. : .....

Name : .....

Third Semester M.A. Degree Examination, June 2022

Behavioural Economics and Data Science

BEDS 534.2 : DATA ANALYTICS FOR BUSINESS

(2020 Admission)

Time : 3 Hours Max. Marks : 75

PART – I

Answer **all** questions in **1 word** to maximum of **2** sentences. Each question carries **1** mark:

1. Regression
2. Unsupervised learning
3. ROC curve
4. Underfitting
5. Text mining
6. Machine learning
7. Nonlinear function
8. Support vectors
9. Clustering
10. Neural network.

(10 × 1 = 10 Marks)

P.T.O.

PART – II

Answer **any seven** questions. Each should not exceed **400** words

11. What is cross validation?

12. Define joint probability distribution.

13. Explain N-gram sequence in text mining.

14. What is a lift curve?

15. Define over fitting.

16. Why do we need data analytics?

17. What is meant by Big Data?

18. Briefly explain regularization.

19. What is meant by baseline in machine learning?

20. What do you mean by decision tree pruning?

($7 \times 5 = 35$ Marks)

PART – III

Answer **any three** questions. Each answer should not exceed **1200** words.

21. What is data mining? What are the steps involved in data mining?

22. What is meant by conditional probability? How it is useful in data analytics?

23. What is clustering? Explain the difference between centroids and connectivity models.

24. Briefly discuss the bag of words model.

25. Define a linear regression model estimated using least square method. Explain the below mentioned linear regression model and interpret the result:

We have some data collected from some car manufacturers. In the model. we try to estimate the impact of Engine power (Hp) and Weight(Wt) of a vehicle on its milage (km/litre). Assume 5% significance

	Estimate	P value
Intercept	29.39	5.13e-05 ***
Hp	–0.032230	0.001178***
wt	-3.227954	0.000364***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R^2 = 0.865 Adjusted R-squared: 0.8194

F-statistic: 47.88 on 3 and 28 DF, p-value: 3.768e-11

Residual standard error: 2.561 on 28 degrees of freedom

Multiple R-squared: 0.8369. Adjusted R-squared: 0.8194

F-statistic: 47.88 on 3 and 28 DF, p-value: 3.768e-11

(3 × 10 = 30 Marks)

———————